



**Pedro Emanuel dos
Santos Santiago**

Visualização de Fluxos Migratórios

Projeto apresentado à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Sistemas de Informação, realizado sob a orientação científica do Dr.-Ing. Joaquim João Estrela Ribeiro Silvestre Madeira, Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática.

o júri

Presidente

Professor Doutor José Manuel Matos Moreira
Professor Auxiliar, Universidade de Aveiro

Arguente principal

Professor Doutor Telmo Eduardo Miranda Castelão da Silva
Professor Auxiliar, Universidade de Aveiro

Orientador

Professor Dr.-Ing. Joaquim João Estrela Ribeiro Silvestre Madeira
Professor Auxiliar, Universidade de Aveiro

agradecimentos

Um agradecimento ao Professor Joaquim Madeira pela orientação, partilha de conhecimentos, opiniões e críticas e no esclarecimento de dúvidas deste projeto.

A todos os que, direta ou indiretamente, contribuíram para o desenvolvimento deste projeto e fizeram parte da minha formação.

Agradeço em especial à minha família pois sem eles seria impossível chegar aqui. Quero agradecer a todos os professores pelo que me ensinaram ao longo do meu percurso académico.

palavras-chave

Visualização de informação, Fluxos migratórios, Plataforma Web, Censos de Portugal

resumo

A importância e quantidade da Informação criada em diversas áreas tornam a sua visualização e compreensão numas das tarefas mais importantes nos dias de hoje. A utilização da visualização de informação para representar vários tipos de dados de uma forma mais acessível tem-se difundido em várias áreas.

Este projeto tem como objetivo representar fluxos migratórios com base em dados fornecidos pelo Instituto Nacional de Estatística. As visualizações geradas servem para facilitar a sua compressão e análise.

Este relatório apresenta uma breve introdução à visualização de informação, apresentando alguns exemplos de aplicações e algumas ferramentas usadas nesta área.

De seguida apresentam-se as visualizações desenvolvidas e suas características, a plataforma que as integra e os resultados de uma primeira avaliação.

keywords

Information visualization, Migration flows, Web platform, Portuguese Census

Abstract

The importance and the amount of information generated by many areas make its visualization and understanding one of the most important tasks of the present. The use of information visualization, to represent in a simple manner several types of data, has spread in many areas.

This project has as objective the representation of migratory flows given the data shared by Statistics Portugal. The generated visualizations are meant to simplify their analysis and understanding.

This project presents a brief introduction to information visualization, some application examples and some common tools used in the area.

Afterwards, the developed visualizations and their features are presented, as well as the integrating platform and the results of a first evaluation.

Índice

| | |
|--|----|
| 1. Introdução..... | 1 |
| 1.1 Motivação e Contexto..... | 1 |
| 1.2 Objetivos | 1 |
| 1.3 Estrutura deste Relatório | 2 |
| 2. Enquadramento..... | 3 |
| 2.1 Visualização de Informação | 3 |
| 2.2 Visualização de Informação versus Visualização de dados..... | 4 |
| 2.3 Processo de Visualização de Informação | 6 |
| 2.4 Visualização de Dados Migratórios..... | 6 |
| 2.5 Ferramentas e Tecnologias | 14 |
| 3. Visualização de Dados dos Censos | 21 |
| 3.1 Conjunto de Dados | 21 |
| 3.2 Informação Disponibilizada dos Censos 2001 e 2011 | 22 |
| 3.3 Processamento dos Dados | 27 |
| 4. Visualizações Desenvolvidas | 29 |
| 4.1 Introdução..... | 29 |
| 4.2 Objetivos e Requisitos Funcionais | 30 |
| 4.3 Visualizações Geradas..... | 30 |
| 4.4.1 Diagrama de Sankey..... | 30 |
| 4.4.2 Diagrama Cordal | 33 |
| 4.4 Operações e Interação..... | 36 |
| 4.5 Resultados | 38 |
| 5. Plataforma | 41 |

| | |
|--|----|
| 5.1 Atores | 42 |
| 5.2 Modelo de Casos de Uso | 42 |
| 5.3 Estrutura do Sistema..... | 43 |
| 5.4 Estrutura de Ficheiros..... | 45 |
| 6. Avaliação..... | 47 |
| 6.1 Avaliação Heurística | 47 |
| 6.2 Resultados | 50 |
| 7. Conclusão e Trabalho Futuro | 53 |
| 8. Referências | 55 |
| Anexo A Organização dos ficheiros..... | 57 |
| Anexo B Manual de Instalação..... | 59 |

Lista de Figuras

| | |
|--|----|
| Figura 2.1 Mapa Lugano-Pisa com o trajeto de forma textual e gráfica [6]. | 4 |
| Figura 2.2 Mapa de Charles Minard da desastrosa Campanha Russa de 1812 de Napoleão. | 5 |
| Figura 2.3 Pipeline de visualização de informação baseado em S. Liu et al [8]. | 6 |
| Figura 2.4 Grafo e Matriz de adjacência [9]. | 7 |
| Figura 2.5 O Mapa de 1858 das migrações do mundo | 9 |
| Figura 2.6 Migração na Inglaterra de pessoas com 60 anos ou mais de Law & Warnes [10] | 10 |
| Figura 2.7 Mapa de migração do mundo | 11 |
| Figura 2.8 Diagrama de Sankey de exportações | 12 |
| Figura 2.9 Excerto do diagrama de Sankey de fluxos migratórios | 13 |
| Figura 2.10 Diagrama cordal de migrações | 14 |
| Figura 2.11 Mapa desenvolvido com o Google Charts | 15 |
| Figura 2.12 Gráfico de barras usando Chart.js | 16 |
| Figura 2.13 Pie Chart usando o Flot Charts [15] | 16 |
| Figura 2.14 Gráfico de barras, séries, utilizando a biblioteca Dc.js [16]. | 17 |
| Figura 2.15 Gráfico utilizando a biblioteca Epoch com duas fontes de dados | 18 |
| Figura 2.16 Plataforma feita em D3.JS. | 18 |
| Figura 3.1 Excerto dos dados utilizados neste trabalho. | 22 |
| Figura 3.2 Modelo lógico da base de dados | 28 |
| Figura 4.1 Exemplo de deslocação da residência para o trabalho | 29 |
| Figura 4.2 Diagrama de Sankey que representa as deslocações das pessoas da residência para o trabalho | 31 |
| Figura 4.3 Diagrama de Sankey de uma zona residência para trabalho representando as deslocações para a região do Grande Porto. | 32 |
| Figura 4.4 Diagrama de Sankey que mostra as alterações de residência relativamente ao ano de 2001(atual, há 1 ano e há 5 anos | 32 |
| Figura 4.5 Diagrama Cordal representando a residência atual e há 1 antes utilizando as NUTS3 | 33 |
| Figura 4.6 Cores dos arcos associados à região da Grande Lisboa. Visualização da residência atual e há um ano | 34 |
| Figura 4.7 Espessura dos arcos associados à região da Grande Lisboa. | 35 |
| Figura 4.8 Filtragem de dados. | 36 |
| Figura 4.9 Diagrama de Sankey com filtragem de dados. | 36 |
| Figura 4.10 Tooltip de informação no diagrama de Sankey. | 37 |
| Figura 4.11 Tooltip do diagrama cordal | 38 |
| Figura 5.1 Plataforma de visualização de fluxos migratórios. | 41 |
| Figura 5.2 Modelo de casos de uso. | 42 |
| Figura 5.3 Modelo MVC | 43 |
| Figura 5.4 Diagrama de componentes | 44 |
| Figura 5.5 Links e nodes do diagrama de Sankey | 45 |

| | |
|---|----|
| Figura 5.6 Exemplo de uma matriz do diagrama cordal..... | 46 |
| Figura 5.7 Excerto dos atributos da matriz do diagrama cordal | 46 |

Lista de tabelas

| | |
|--|----|
| Tabela 2.1 Representação do fluxo de migrações com base em origem-destino em forma de matriz adjacência..... | 7 |
| Tabela 2.2 Lista de adjacência da amostra de dados do projeto | 8 |
| Tabela 2.3 Resumo de algumas características das tecnologias abordadas | 19 |
| Tabela 3.1 Descrição dos dados utilizados..... | 23 |
| Tabela 3.2 Descrição dos atributos da tabela meio vida..... | 23 |
| Tabela 3.3 Excerto dos dados da tabela meio de vida | 24 |
| Tabela 3.4 Descrição dos atributos do sexo..... | 24 |
| Tabela 3.5 Atributos do sexo..... | 24 |
| Tabela 3.6 Descrição dos atributos do alfabeto | 24 |
| Tabela 3.7 Atributos do alfabeto | 24 |
| Tabela 3.8 Descrição dos atributos das NUTS2 | 25 |
| Tabela 3.9 Atributos das NUTS2 | 25 |
| Tabela 3.10 Descrição dos atributos das NUTS3 | 25 |
| Tabela 3.11 Atributos das NUTS3 | 26 |

1. Introdução

1.1 Motivação e Contexto

Grande parte da informação está armazenada digitalmente. Estima-se que em 2010 tenha sido armazenado cerca de 1,200 *exabytes* de dados [1].

Este volume de dados contém informações valiosas desde informações das empresas, associações, estado e vida pessoal, proporcionando assim um melhoramento para o dia a dia da sociedade. No entanto, para tomar decisões adequadas, temos de encontrar maneiras de explorar, relacionar e comunicar os dados de modo inteligente [1].

A visualização de informação tem como objetivo principal ajudar as pessoas na compreensão dos dados, aproveitando assim a capacidade do sistema visual humano para ver padrões, identificar tendências e valores diferentes [1].

Este projeto está inserido na área de visualização de informação e pretende apresentar várias formas de visualizar as migrações de pessoas recenseadas presentes na amostra de dados dos censos de 2001 e 2011. A escolha deste tema deve-se ao facto de ter interesse nesta área, ao grande volume de dados e ao interesse familiar também na área geográfica.

1.2 Objetivos

O objetivo principal deste projeto é desenvolver visualizações e uma plataforma que permita visualizar os movimentos migratórios, permitindo assim uma melhor análise dos dados utilizados neste projeto.

Esta plataforma deve permitir visualizar fluxos migratórios de uma determinada região para outras e ainda analisar as entradas e saídas.

As visualizações devem ser integradas numa plataforma que seja fácil de utilizar para os utilizadores.

1.3 Estrutura deste Relatório

No capítulo 2 é apresentada uma abordagem à visualização de informação, são introduzidos os dados migratórios e mostradas algumas aplicações já existentes nesta área e, por fim, são apresentadas as ferramentas que podem ser utilizadas.

No capítulo 3 são introduzidos e analisados os dados das amostras dos censos. São apresentados ainda alguns exemplos dos dados originais e o processamento realizado sobre os mesmos.

No capítulo 4 são apresentados os objetivos das visualizações, os requisitos funcionais e quais foram as visualizações geradas incluindo as suas operações e interações.

No capítulo 5 é apresentada a plataforma implementada e casos de uso.

No capítulo 6 são apresentados os resultados de uma avaliação heurística à plataforma.

No capítulo 7 são apresentadas as conclusões deste projeto e algumas ideias para trabalho futuro.

2. Enquadramento

Neste capítulo é realizado um enquadramento teórico sobre o tema deste projeto. Está dividido em duas partes, numa primeira é feita uma contextualização sobre visualização de informação e numa segunda parte é realizada uma introdução às visualizações de dados migratórios mostrando exemplos de aplicação.

2.1 Visualização de Informação

“Information visualization is a research area that aims to aid users in exploring, understanding, and analyzing data through progressive, iterative visual exploration” [2].

Visualização de informação (*InfoVis – Information Visualization*) é uma área de estudo aplicada em várias áreas como por exemplo: na meteorologia, medicina, biologia, indústria, etc.

É um processo de mapeamento de dados, abrangendo representações visuais e técnicas de interação. Tem como finalidade representar dados que podem ser abstratos ou/e não espaciais [3].

Através da capacidade visual humana e da perceção [4], torna-se mais fácil compreender uma imagem acompanhada com texto [5]. Na figura 2.1 é apresentado o trajeto Lugano-Pisa com os detalhes do lado esquerdo. Com a utilização da cor no mapa consegue-se identificar rapidamente o trajeto proposto, bem como estradas que podem servir de alternativa [6].



Figura 2.1 Mapa Lugano-Pisa com o trajeto de forma textual e gráfica [6].

2.2 Visualização de Informação versus Visualização de dados

A visualização distingue-se em dois tipos, visualização de informação e visualização de dados. Chen ajuda-nos a compreender a diferença entre estas duas abordagens [3].

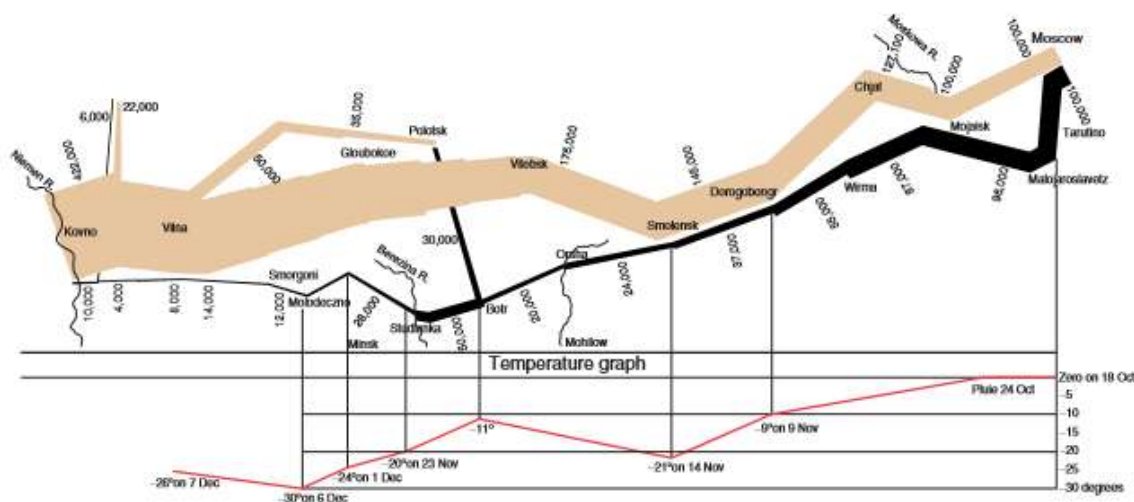


Figura 2.2 Mapa de Charles Minard da desastrosa Campanha Russa de 1812 de Napoleão¹

O mapa (Figura 2.2) criado por Charles Minard representa o percurso do exército de Napoleão, a faixa maior mostra o tamanho do exército em locais específicos durante o seu avanço e recuo. O mapa exibe seis dados e duas dimensões, o número de tropas de Napoleão; a distância percorrida; temperatura; latitude e longitude; direção da viagem; e localização em relação a datas específicas [3].

A evolução da perda de vidas está visível nos locais onde teve maior impacto:

- Dos 422 000 soldados que partiram apenas 10 000 voltaram com vida
- Um fator importante para as mortes foi a temperatura.

Esta visualização pode ser considerada um exemplo do que hoje em dia se chama de diagrama de Sankey [4].

O que Minard disse sobre a sua visualização:

“The aim of my carte figurative is less to express statistical results, better done by numbers, than to convey promptly to the eye the relation not given quickly by numbers requiring mental calculation” [7].

A diferença entre visualização de informação e visualização de dados pode ser definida como a presença de dados quantitativos ou não e o quão fácil se pode transformá-los em formas quantitativas [3].

¹ <http://waterloo200.org/200-object/flow-map-of-napoleons-invasion-of-russia/> (consultado em 26-09-2016)

2.3 Processo de Visualização de Informação

Na figura 2.3 é representado o *pipeline* de visualização de informação que se encontra dividido em cinco fases:

- **Transformação de dados e análise:** este passo consiste em analisar e transformar os dados de forma a facilitar as seguintes fases do processo.
- **Filtragem:** esta etapa permite seleccionar os dados que se pretende visualizar.
- **Mapeamento de Informação:** depois de filtrados os dados estes são mapeados em primitivas de visualização (e.g. pontos, polígonos) e nos atributos dessas primitivas (e.g. cor, tamanho, posição).
- **Representação das primitivas gráficas:** esta fase é a responsável por representar as primitivas e gerar a representação dos dados.
- **Interação com o utilizador:** os utilizadores poderão interagir com os controlos disponíveis.

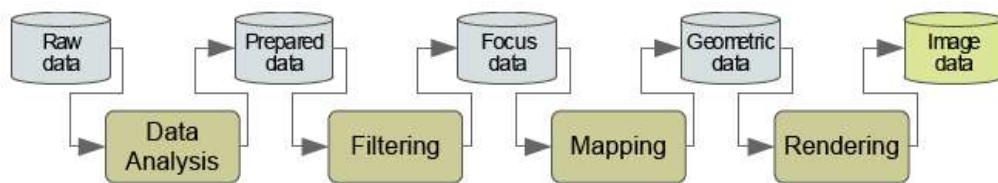


Figura 2.3 Pipeline de visualização de informação baseado em S. Liu et al [8]

2.4 Visualização de Dados Migratórios

Os dados migratórios nada têm a ver com mortalidade ou fertilidade. A migração ocorre devido a vários fatores que podem ser a nível económico, social e tecnológico.

Os dados migratórios estão relacionados com entrada e saída de pessoas de uma determinada zona para outra diferente. Quando os movimentos ocorrem dentro do mesmo país chamam-se migrações internas.

Os dados do fluxo migratório representam relações geográficas, nos quais se especificam as origens, destino e as quantidades. Na figura 2.4 é representada uma matriz de adjacência e o grafo resultante [9].

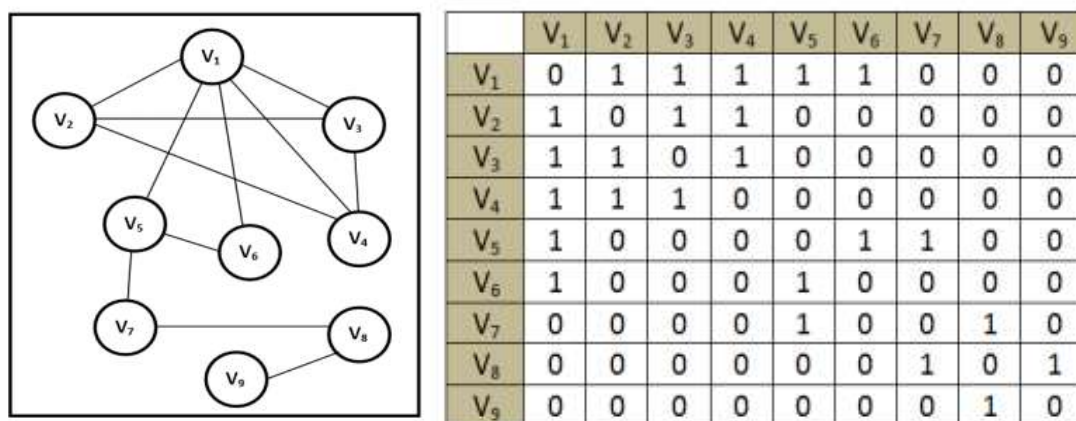


Figura 2.4 Grafo e Matriz de adjacência [9]

Os dados de fluxo migratório são frequentemente representados e armazenados numa matriz. Na tabela 2.1 está representada uma matriz com linhas e colunas referente à origem e ao destino, cujos valores dizem respeito às regiões de Portugal (NUTS2). Os valores contidos nas células representam a quantidade de pessoas que se deslocam de uma determinada zona para outra.

| Origem/Destino | Norte | Centro | Lisboa | Algarve | Alentejo | Madeira | Açores |
|-----------------|--------|--------|--------|---------|----------|---------|--------|
| Norte | 110042 | 1533 | 810 | 148 | 180 | 71 | 66 |
| Centro | 1211 | 63709 | 2249 | 143 | 617 | 34 | 39 |
| Lisboa | 279 | 960 | 81247 | 272 | 726 | 54 | 46 |
| Algarve | 37 | 143 | 240 | 11202 | 95 | 2 | 4 |
| Alentejo | 46 | 617 | 1265 | 195 | 19542 | 8 | 8 |
| Madeira | 48 | 53 | 92 | 6 | 17 | 7291 | 7 |
| Açores | 33 | 42 | 132 | 8 | 11 | 15 | 6838 |

Tabela 2.1 Representação do fluxo de migrações com base em origem-destino em forma de matriz adjacência.

Noutras visualizações torna-se mais fácil utilizar uma representação na forma de lista de adjacência, como por exemplo no diagrama de Sankey. A cada arco podem ser associados vários atributos (ver tabela 2.2).

| Origem | Destino | Atributo 1 | Atributo 2 | ... |
|--------------|----------|------------|------------|-----|
| Norte | Norte | 110042 | Vermelho | ... |
| Norte | Algarve | 143 | Vermelho | ... |
| Norte | Centro | 1533 | Vermelho | ... |
| Norte | Lisboa | 81 | Vermelho | ... |
| Norte | Alentejo | 180 | Vermelho | ... |
| ... | ... | ... | ... | ... |

Tabela 2.2 Lista de adjacência da amostra de dados do projeto

Exemplos de Aplicação

Ao longo dos anos as visualizações de fluxos migratórios foram evoluindo. De seguida são representados alguns exemplos dessas visualizações.

Na figura 2.5 é representado o mapa de Minard² que mostra os fluxos migratórios com base no ano de 1858. Este mapa ainda hoje é conhecido na área de Visualização de Informação. O mapa desenhado à mão contém as origens, destinos e a magnitude dos fluxos em 1858, sendo que um milímetro corresponde 1500 pessoas. O mapa de Minard é visualmente cativante pois mostra apenas um pequeno número de fluxos [10].

Naquele ano, 86.000 ingleses, 45.300 alemães, 20.000 franceses e 11.600 portugueses, deixaram os seus países de origem.

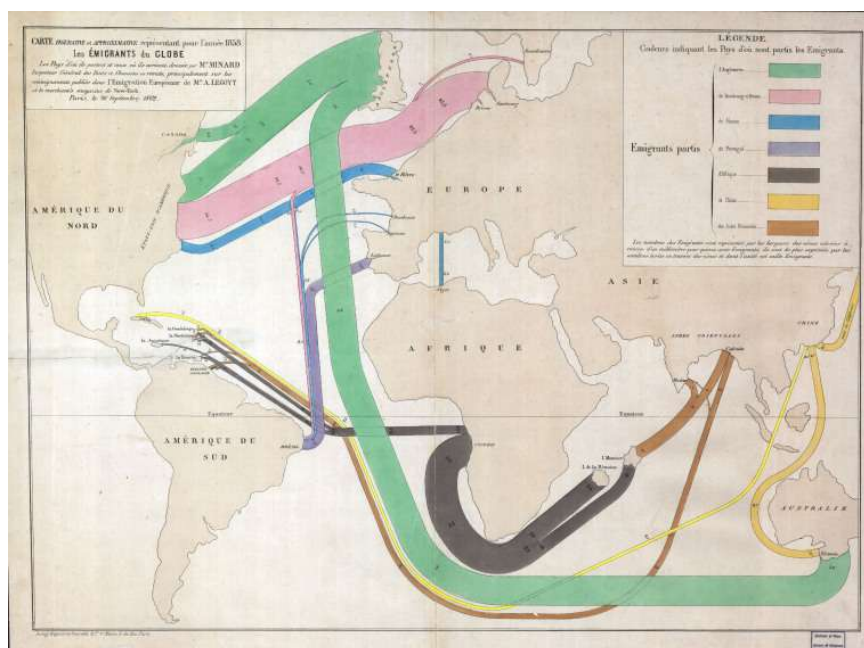


Figura 2.5 O Mapa de 1858 das migrações do mundo³

Uma desvantagem deste tipo de mapa é que funcionam bem apenas para um pequeno conjunto de dados [10].

² <http://cartographia.wordpress.com/category/charles-joseph-minard> (consultado a 26-09-2016)

³ <http://www.citylab.com/work/2014/11/international-migration-patterns-from-1858/382524/> (consultado a 26-09-2016)

A utilização de um maior volume de dados é representada na figura 2.6, um exemplo de um mapa em que as setas indicam a direção e o valor do fluxo. Este mapa representa a migração de pessoas na Inglaterra com 60 ou mais anos entre os anos 1961 a 1966. É notório o domínio de Londres da população mais idosa na Inglaterra [10].

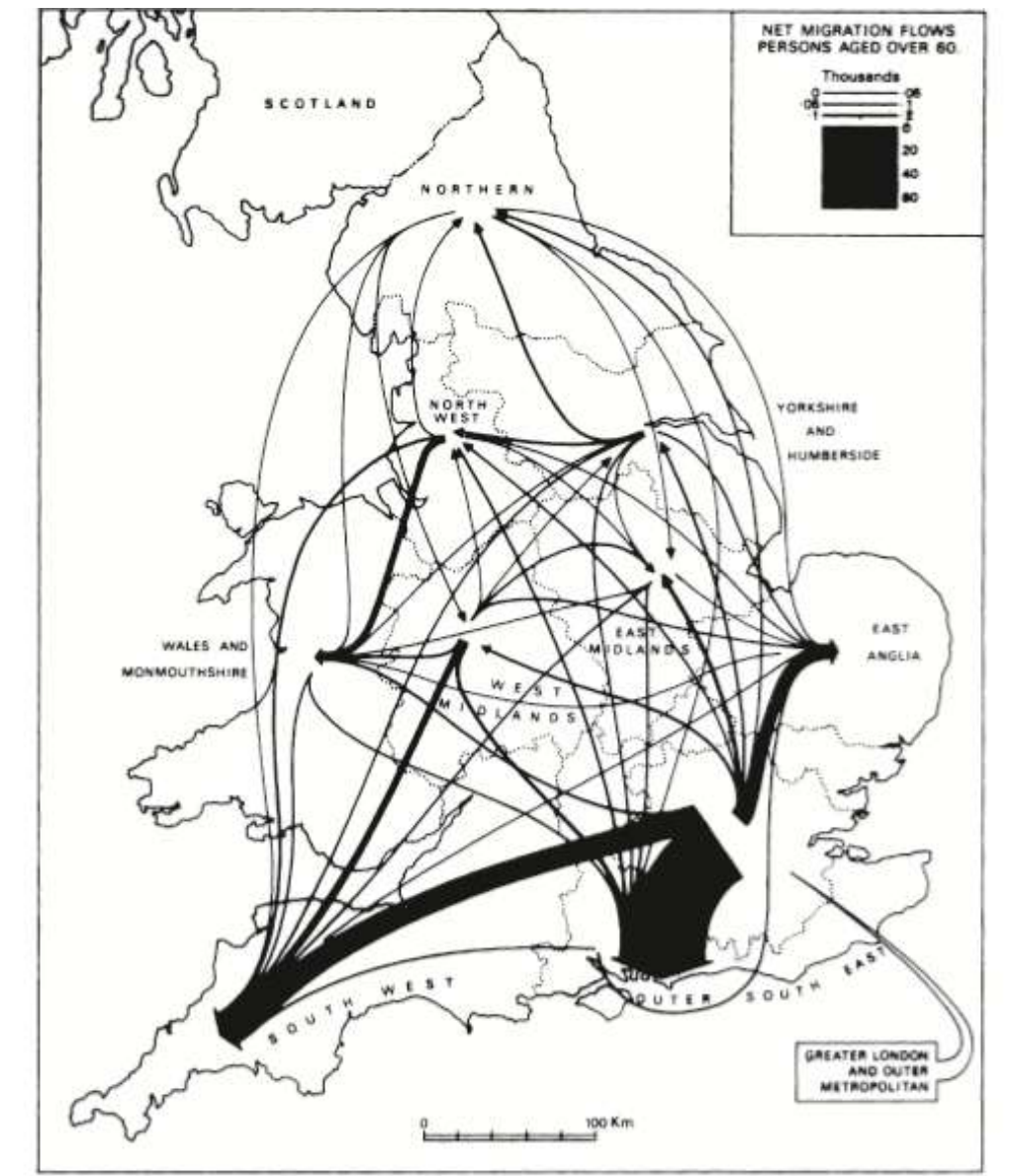


Figura 2.6 Migração na Inglaterra de pessoas com 60 anos ou mais de Law & Warnes [10]

Grande parte das visualizações de hoje em dia são desenvolvidas para a *Web*. Na figura 2.7 é representado o mapa do mundo onde estão representados os fluxos migratórios. Podemos ainda visualizar informação adicional tal como: total da população, PIB, HIV, taxa de mortalidade de crianças menores de cinco anos.

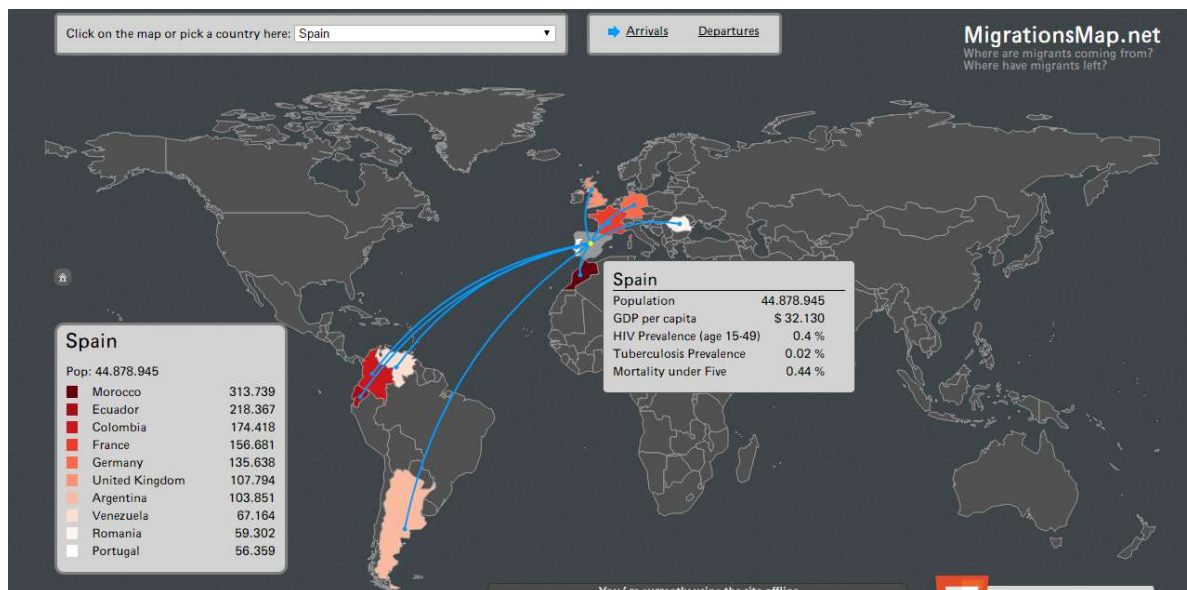


Figura 2.7 Mapa de migração do mundo ⁴

Diagrama de Sankey

O diagrama de Sankey foi inicialmente desenvolvido por Matthew Sankey⁵ e é normalmente utilizado para visualizar fluxos de energia ou materiais. Representa informação quantitativa sobre o fluxo de dados e movimentos. A espessura de cada ligação ou arco é proporcional à quantidade transferida [7].

Com o diagrama de Sankey, consegue-se perceber, à primeira vista, as origens e destinos com maior importância e compreender as ligações com maior fluxo entre determinadas origens e destinos.

Na figura 2.8 é representado o diagrama de Sankey, que mostra os fluxos de exportação de cavalos na União Europeia.

⁴ <http://migrationsmap.net/> (consultado a 20-10-2016)

⁵ <http://www.sankey-diagrams.com/who-is-this-sankey-guy/> (consultado a 26-09-2016)

Horse Exports *from* EU country *to* EU country

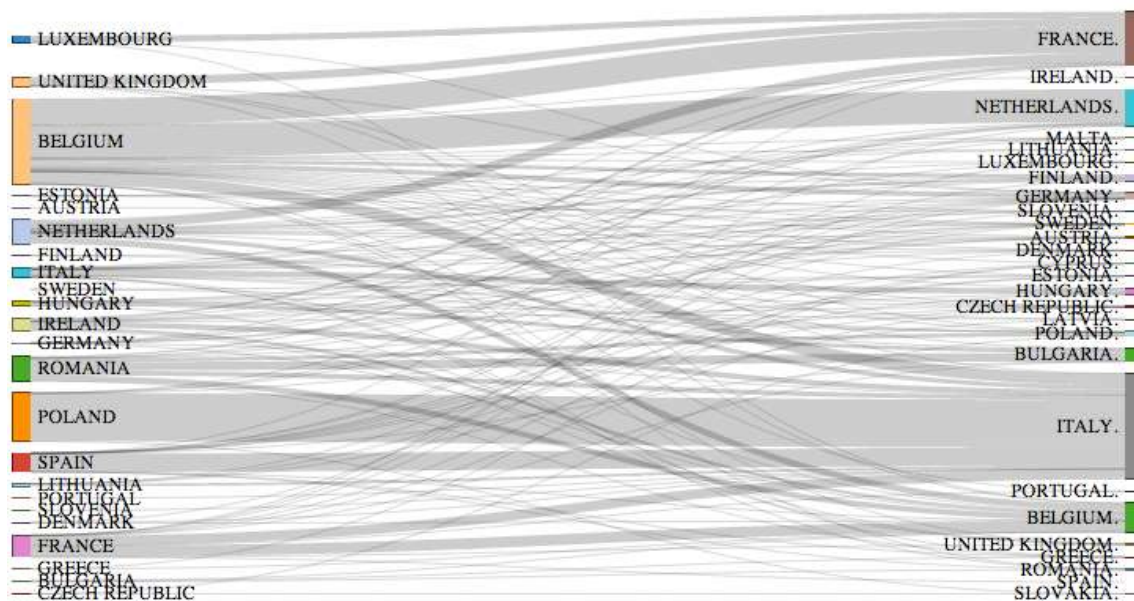


Figura 2.8 Diagrama de Sankey de exportações ⁶

É apresentado na figura 2.9 o digrama de Sankey criado por Carlo Zapponi que representa o fluxo migratório de pessoas e está dividido em duas colunas, à esquerda estão os países de origem e à direita os países de destino.

Utiliza o HTML5 para criar animações personalizadas e a criação dos fluxos. Nesta visualização, o utilizador pode ainda seleccionar apenas um país e visualizar as suas ligações, tendo assim uma perceção mais detalhada da informação.

Esta representação apresenta um problema de usabilidade, uma vez que não possível visualizar todos os fluxos de entrada e saída.

⁶<https://blog.ouseful.info/2013/02/18/reshaping-horse-importexport-data-to-fit-a-sankey-diagram> (consultado a 26-09-2016)

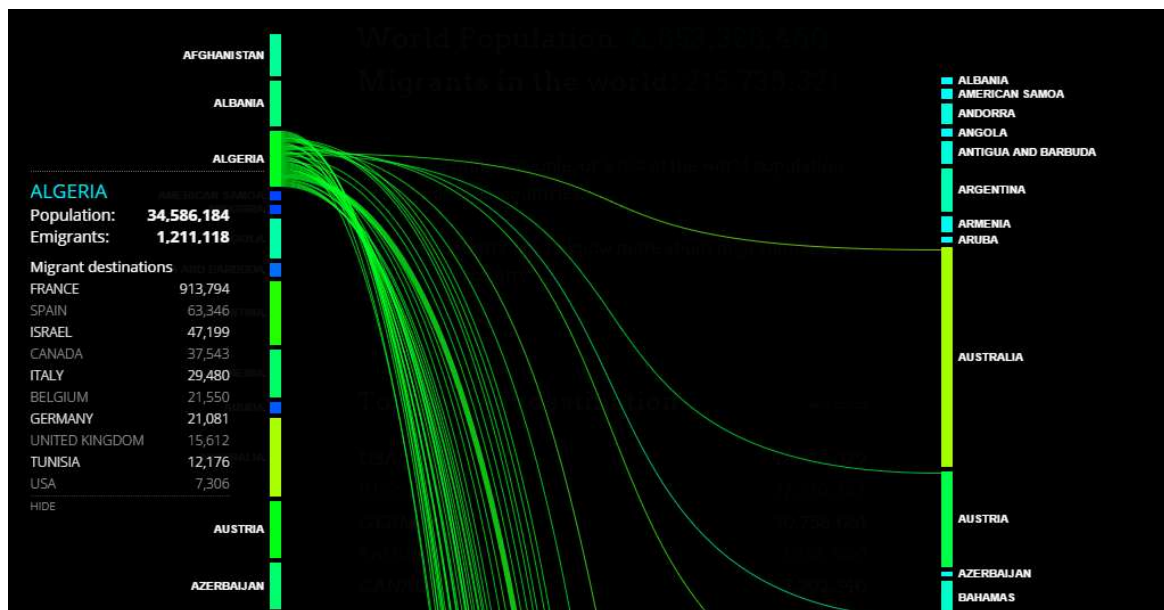


Figura 2.9 Excerto do diagrama de Sankey de fluxos migratórios

Diagrama Cordal

Os diagramas cordais utilizam um *layout* circular para representar locais. As ligações entre origem-destino são representadas por arcos de tamanho variável que ligam dois nós. Através dos arcos consegue-se perceber o volume do fluxo em ambas direções, o tamanho é diferente em cada arco, o que não acontece com o diagrama de Sankey.

Na figura 2.10, o diagrama de Nikola Sander e da sua equipa representa o fluxo de migrações entre 196 países. Os dados representam fluxos migratórios entre 1990 a 2010 com intervalos de 5 anos [10].



Figura 2.10 Diagrama cordal de migrações ⁷

Este tipo de visualização circular é interessante porque permite obter facilmente informações sobre o fluxo de migrações.

Além destas representações existem muitos mais exemplos que podem ser consultadas nos mais recentes *surveys* [11] [12] [5].

2.5 Ferramentas e Tecnologias

Para visualização de informação existem várias bibliotecas que podem ser utilizadas para o desenvolvimento de aplicações gráficas, a nível *mobile* e *Web*. Porém, a maioria das visualizações, hoje em dia, é mais virada para a *Web*.

No desenvolvimento deste projeto a utilização de uma ferramenta direcionada para *Web* é mais vantajosa permitindo assim abranger um público mais alargado. Algumas das

⁷ <http://www.global-migration.info/> (consultado a 26-09-2016)

bibliotecas que poderiam ser utilizadas no âmbito deste projeto são: *Google Charts*, *ChartJS*, *Flot Charts*, *Dc*, *Epoch*, *D3js*.

Google Charts

O *Google Charts*⁸ é uma ferramenta que permite a criação de gráficos e tem uma diversidade de exemplos disponíveis que podem ajudar no seu desenvolvimento.

A biblioteca possibilita o tratamento de dados, desde ordenação de dados, filtragem e modificação. Possibilita alterar a aparência dos gráficos adicionando títulos, cores, alterando a largura e a altura.

O *Google Charts* corre em todos os *browsers* e tem uma licença gratuita para o uso particular.

Na figura 2.11 está representado o mapa com a informação populacional de alguns países, criado com o Google Charts [13].

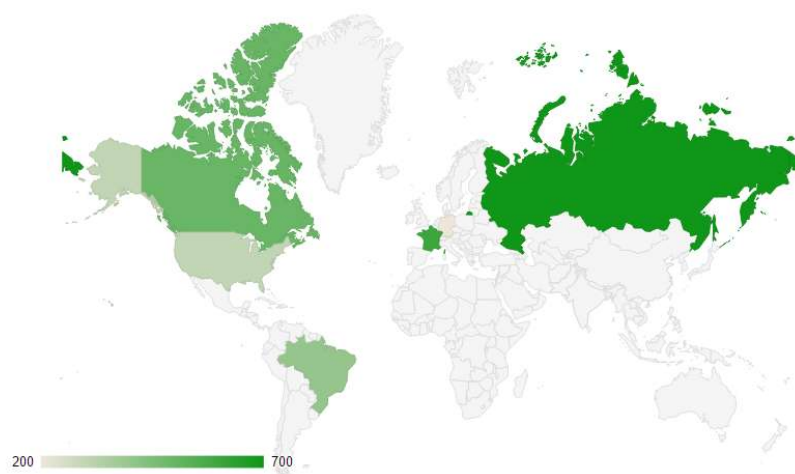


Figura 2.11 Mapa desenvolvido com o Google Charts

⁸ <https://developers.google.com/chart/interactive/docs/gallery/geochart> (consultado a 23-09-2016)

Chart JS

O Charts.js⁹ é uma biblioteca que permite criar facilmente ambientes animados e gráficos interativos para se utilizar em páginas Web. O Charts.js permite criar até 6 tipos de visualizações e é compatível com a maioria dos *browsers*.

A biblioteca é totalmente grátis e tem como principal vantagem as visualizações criadas serem responsivas por omissão, ao invés de outras bibliotecas [14].

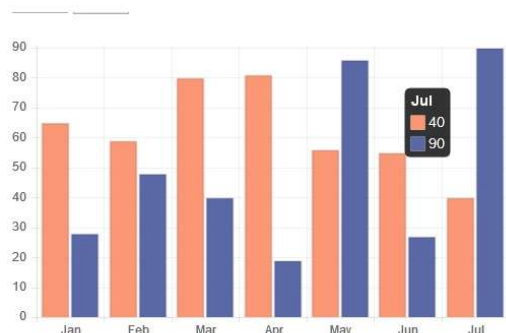


Figura 2.12 Gráfico de barras usando Chart.js

Flot Charts

*Flot Charts*¹⁰ é uma biblioteca para criação de gráficos interativos. O *Flot Charts* tem uma grande variedade de gráficos na sua galeria, bem como código fonte. Utiliza *Jquery* para a desenvolvimento de visualizações e funciona com os seguintes dos *browsers*: Internet Explorer 6+, Chrome, Firefox 2+, Safari 3+ e Opera 9.5+ [15].

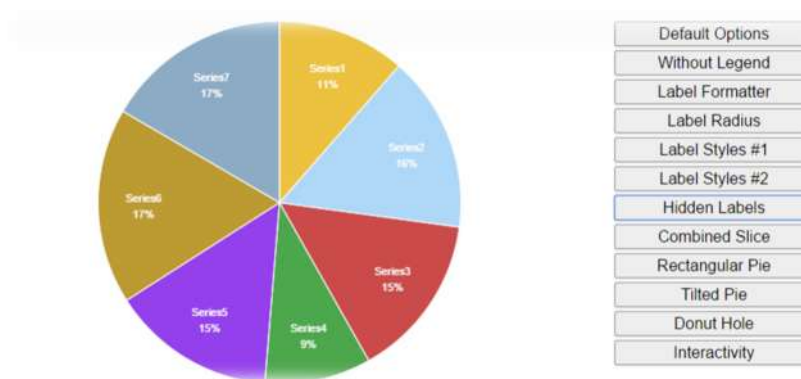


Figura 2.13 Pie Chart usando o Flot Charts [15]

⁹ <http://www.chartjs.org/> (consultado a 23-09-2016)

¹⁰ <http://www.flotcharts.org/> (consultado a 23-09-2016)

Dc.js

O Dc.js¹¹ é uma biblioteca para criação de gráficos em *javascript* que utiliza o *crossfilter*¹², permitindo uma exploração eficiente de grandes conjuntos de dados.

A biblioteca é fácil de utilizar para visualização permitindo ao utilizador interagir com vários gráficos ao mesmo tempo, como se pode visualizar no próprio site de apoio à biblioteca.

Esta biblioteca pode ser usada em qualquer *browser* e nos dispositivos móveis [16].

Na figura 2.14, o utilizador pode seleccionar uma ou mais zonas do mapa e são atualizados os valores dos gráficos. Este gráfico representa as diferenças das taxas de criminalidade entre as principais cidades do Canadá.

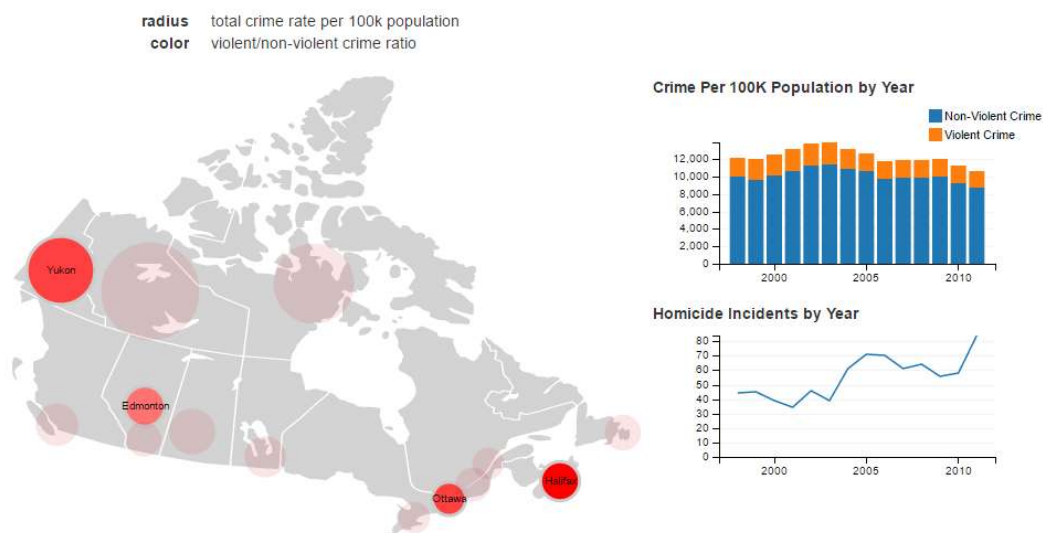


Figura 2.14 Gráfico de barras, séries, utilizando a biblioteca Dc.js [16]

Epoch

A biblioteca *Epoch* foi desenvolvida por Ryan Sandor Richards. Os gráficos *Epoch* são implementados usando o D3. Permite a criação de gráficos básicos que são utilizados na representação de relatórios históricos e gráficos em tempo real que servem para representar dados que são atualizados ao longo do tempo.

¹¹ <https://dc-js.github.io/dc.js/> (consultado a 23-09-2016)

¹² <http://square.github.io/crossfilter/> (consultado a 23-09-2016)

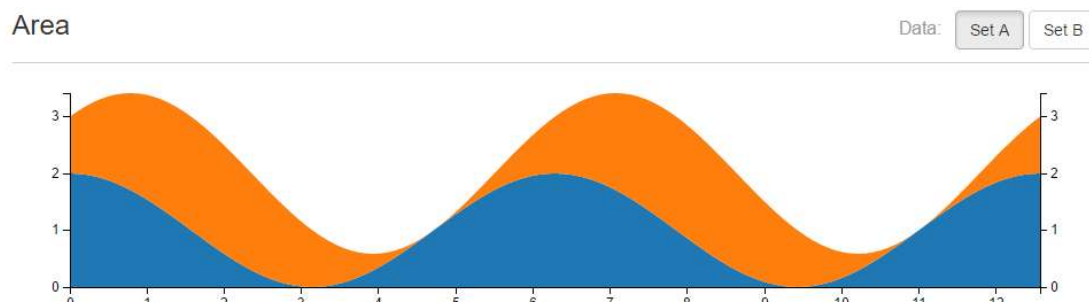


Figura 2.15 Gráfico utilizando a biblioteca Epoch com duas fontes de dados ¹³

D3.JS

O D3.JS foi desenvolvido por Michael Bostock e é uma ferramenta para o desenvolvimento de visualizações avançadas [17].

Esta biblioteca utiliza as seguintes linguagens, HTML, CSS, *JavaScript* e elementos SVG.

O D3.JS permite uma manipulação de dados tendo as suas próprias funções de leitura e processamento. Além disso, o D3.JS permite manipular diretamente elementos da base de dados, combinando componentes de visualização com a manipulação de documentos DOM (*document object model*).

A principal vantagem do D3.JS é que permite representar uma grande quantidade de informação.

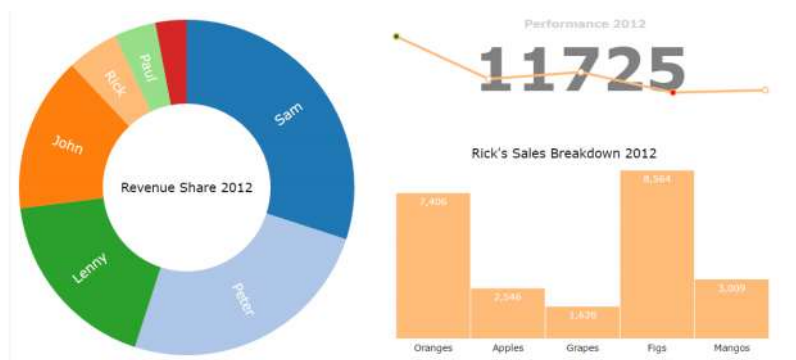


Figura 2.16 Plataforma feita em D3.JS¹⁴

¹³ <https://epochjs.github.io/epoch/basic/#area> (consultado a 23-09-2016)

¹⁴ <http://bl.ocks.org/diethardsteiner/3287802> (consultado 23-09-2016)

Na tabela 2.3 é apresentado um pequeno resumo sobre as tecnologias abordadas anteriormente.

| Características /Ferramentas | Google Charts | Charts JS | Flot Charts | Dc | Epoch | D3 |
|--|------------------|-----------|-------------|-----|-------|-----|
| Grande volume de dados | Sim | Não | Sim | Sim | Não | Sim |
| Visualizações Responsivas | Não | Sim | Não | Sim | Não | Não |
| Compatível com a maioria dos browsers | Sim | Sim | Não | Sim | Sim | Sim |
| Muitos tipos de gráficos | Sim | Sim | Não | Não | Não | Sim |
| Próprias funções de leitura de dados | Sim | Não | Não | Não | Não | Sim |
| Interatividade com as visualizações | Sim | Sim | Sim | Sim | Sim | Sim |
| Documentação de apoio | Sim | Sim | Não | Sim | Não | Não |

Tabela 2.3 Resumo de algumas características das tecnologias abordadas

3. Visualização de Dados dos Censos

Os censos permitem obter um conjunto de informações importantes tais como: total de homens, mulheres, crianças e idosos; permitem, ainda obter informações relativas às deslocações das pessoas entre os locais de nascimento, residência e trabalho [18].

Neste capítulo são apresentadas as amostras de dados dos censos de 2001 e 2011 que foram utilizados para desenvolver as visualizações.

3.1 Conjunto de Dados

Os dados disponibilizados pelo Instituto Nacional de Estatística (INE) são uma amostra de 5% e representam os anos de 2001 e 2011 (estes dados foram retirados do site oficial do Instituto Nacional de Estatística¹⁵).

Entrámos em contacto com o INE para tentar obter mais dados. Só poderiam ser disponibilizados mediante pagamento.

O tamanho da base de dados disponibilizada pelos censos dos dois anos ronda 1.000.000 de registos.

Os dados disponibilizados tinham informação que não era relevante para o âmbito deste projeto, como por exemplo, o alojamento dos habitantes.

Após uma análise cuidadosa dos dados e face às necessidades do projeto, ficaram determinados quais os dados a serem utilizados para construir as visualizações deste trabalho. Na secção a seguir são apresentados os dados utilizados neste projeto.

¹⁵ <https://www.ine.pt/>(consultado a 23-09-2016)

3.2 Informação Disponibilizada dos Censos 2001 e 2011

Nesta amostra é possível obter um aglomerado de informação quantitativa, tais como, o número total de habitantes por concelho (município).

No conjunto de dados desta amostra para 2001 existem 249.813 homens e 267.902 mulheres, num total de 517.715 indivíduos; em 2011, existem 240.475 homens e 262.693 mulheres num total de 503.168 indivíduos, sendo que a grande maioria das pessoas pertencem às regiões do Porto e Lisboa.

Na figura 3.1 está representado um excerto dos dados utilizados neste projeto. Esta tabela contém informação sobre cada indivíduo, isto é, o local onde nasceram, residiram e onde residem atualmente, bem como o sexo, o meio de vida, a idade, etc. O significado de cada coluna está representado na tabela 3.1.

Não são representados os atributos associados às colunas NUTS3_nat, NUTS3_trab, NUTS3_resid_1amc e NUTS3_resid_5amc pois representam a mesma informação da tabela NUTS3 que está representada na tabela 3.11.

| censo | nuts2_eu02 | nuts3_eu02 | sexo | idade | nuts3_nat | nuts3_trab | nuts3_resid_1amc | nuts3_resid_5amc | meiovida | alfabetismo |
|-------|------------|------------|------|-------|-----------|------------|------------------|------------------|----------|-------------|
| 2011 | 17 | 171 | 1 | 38 | 185 | 185 | 185 | 185 | 1 | 1 |
| 2011 | 17 | 171 | 2 | 34 | 171 | 171 | 171 | 171 | 1 | 1 |
| 2011 | 17 | 171 | 2 | 47 | 165 | 165 | 165 | 165 | 1 | 1 |
| 2011 | 17 | 171 | 2 | 11 | 171 | 171 | 171 | 171 | 9 | 1 |
| 2011 | 17 | 171 | 1 | 70 | 999 | 999 | 999 | 999 | 2 | 1 |
| 2011 | 17 | 171 | 2 | 66 | 999 | 999 | 999 | 999 | 2 | 1 |
| 2011 | 17 | 171 | 2 | 64 | 117 | 117 | 117 | 117 | 2 | 1 |
| 2011 | 17 | 171 | 2 | 93 | 117 | 117 | 117 | 117 | 2 | 1 |
| 2011 | 17 | 171 | 2 | 77 | 171 | 171 | 171 | 171 | 2 | 1 |
| 2011 | 17 | 171 | 2 | 62 | 999 | 999 | 999 | 999 | 1 | 1 |

Figura 3.1 Excerto dos dados utilizados neste trabalho

A tabela 3.1 apresenta uma descrição detalhada acerca das variáveis utilizadas neste projeto.

| Nome | Descrição |
|-------------------|--------------------------|
| Censo | Ano do Censo |
| NUTS2_eu02 | NUTS 2 Geografia de 2002 |
| NUTS3_eu02 | NUTS 3 Geografia de 2002 |
| Sexo | Sexo do indivíduo |

| | |
|-------------------------|---|
| NUTS3_Trab | NUTS3 de trabalho ou estudo |
| NUTS3_nat | NUTS3 de naturalidade |
| NUTS3_resid_1amc | NUTS3 de residência anterior (1 anos antes) |
| NUTS3_resid_5amc | NUTS3 de residência anterior (5 anos antes) |
| Idade | Idade do indivíduo |
| Meia vida | Principal meio de vida |
| Alfabetismo | Alfabetismo |

Tabela 3.1 Descrição dos dados utilizados

Na tabela 3.2 temos a descrição da tabela meio de vida, na tabela 3.3 são representados os dados do meio de vida da referida amostra. O meio de vida representa a forma de sobrevivência das pessoas.

| Nome | Descrição |
|------------------|---------------------|
| Meia vida | Código |
| DSG | Descrição do Código |

Tabela 3.2 Descrição dos atributos da tabela meio vida

| MEIOVIDA | DSG |
|------------|--------------------------------------|
| 411 | Trabalho |
| 12 | Subsídio de Doença, Acidente, etc. |
| 13 | Subsídio de Desemprego |
| 14 | Outros Subsídios |
| 15 | Rendimento Mínimo Garantido |
| 16 | Pensão / Reforma |
| 17 | Rendimento de Propriedade ou Empresa |

| | |
|-----------|--------------------|
| 18 | Apoio Social |
| 19 | A cargo da família |
| 20 | Outros Casos |

Tabela 3.3 Excerto dos dados da tabela meio de vida

Na amostra de dados são discriminados os dois géneros, homens e mulheres. Na tabela 3.4 temos a descrição da tabela sexo e, na tabela 3.5 os atributos do sexo.

| Nome | Descrição |
|-------------|---------------------|
| Sexo | Código |
| DSG | Descrição do Código |

Tabela 3.4 Descrição dos atributos do sexo

| SEXO | DSG |
|----------|-----------|
| 1 | Masculino |
| 2 | Feminino |

Tabela 3.5 Atributos do sexo

Em Portugal, a taxa de analfabetismo tem vindo a diminuir ao longo das últimas décadas, devido às melhores condições socioeconómicas. Na tabela 3.6 é representada uma descrição dos dados e na tabela 3.7 são apresentados o conjunto de dados do alfabetismo.

| Nome | Descrição |
|-----------------|---------------------|
| Alfabeto | Código |
| DSG | Descrição do Código |

Tabela 3.6 Descrição dos atributos do alfabeto

| ALFABETO | DSG |
|----------|---------------------------|
| 1 | Sabe ler e escrever |
| 3 | Não sabe ler nem escrever |

Tabela 3.7 Atributos do alfabeto

Portugal está dividido em regiões NUTS 1, 2, 3.

As NUTS1 compreendem as regiões autónomas da Madeira, dos Açores e o Continente.

As NUTS2 são compostas por 7 unidades: 5 no Continente e 2 nos Açores e Madeira.

As NUTS3 são compostas por 25 regiões das quais 23 no Continente e 2 nos Açores e Madeira.

Nas NUTS3 os dois primeiros algarismos correspondem à respetiva NUTS2, como, por exemplo: o código 161 corresponde à zona Centro, ou seja, ao código 16 da tabela NUTS2.

| Nome | Descrição |
|-----------------|---------------------|
| Nuts2 | Código |
| Nut2_DSG | Descrição do Código |

Tabela 3.8 Descrição dos atributos das NUTS2

| Nuts2 | NUT2_DSG |
|-----------|----------------------------|
| 11 | Norte |
| 16 | Centro |
| 17 | Lisboa |
| 18 | Alentejo |
| 15 | Algarve |
| 20 | Região Autónoma dos Açores |
| 30 | Região Autónoma da Madeira |

Tabela 3.9 Atributos das NUTS2

| Nome | Descrição |
|-----------------|---------------------|
| Nuts3 | Código |
| Nut3_DSG | Descrição do Código |

Tabela 3.10 Descrição dos atributos das NUTS3

| NUTS3 | NUT3_DSG |
|------------|------------|
| 111 | Minho-Lima |
| 112 | Cávado |
| 113 | Ave |

| | |
|-------------------|----------------------------|
| 114 | Grande Porto |
| 115 | Tâmega |
| 116 | Entre Douro e Vouga |
| 117 | Douro |
| 118 | Alto Trás-os-Montes |
| 161 | Baixo Vouga |
| 162 | Baixo Mondego |
| 163 | Pinhal Litoral |
| 164 | Pinhal Interior Norte |
| 165 | Dão-Lafões |
| 166 | Pinhal Interior Sul |
| 167 | Serra da Estrela |
| <u>168</u> | Beira Interior Norte |
| 169 | Beira Interior Sul |
| 16A | Cova da Beira |
| 16B | Oeste |
| 16C | Médio Tejo |
| 171 | Grande Lisboa |
| 172 | Península de Setúbal |
| 181 | Alentejo Litoral |
| 182 | Alto Alentejo |
| 183 | Alentejo Central |
| 184 | Baixo Alentejo |
| 185 | Lezíria do Tejo |
| 150 | Algarve |
| 200 | Região Autónoma dos Açores |
| 300 | Região Autónoma da Madeira |

Tabela 3.11 Atributos das NUTS3

3.3 Processamento dos Dados

Os dados estavam disponibilizados em duas bases de dados Access 2001 e 2011. Foi necessário exportar os dados para o *PostgreSQL*¹⁶ utilizando a ferramenta bulzip¹⁷.

Como os dados disponibilizados estavam em duas bases de dados foi necessário centralizar a informação numa única base de dados.

Durante o processo de inserção na base de dados, existiam dados que não estavam relacionados, pelo que foi necessário inserir dados manualmente, por exemplo, regiões sem qualquer tipo de designação. A base de dados criada está representada na figura 3.2.

Para facilitar a modelação da base de dados foi essencial saber o que queríamos visualizar. Foram ainda estabelecidos alguns objetivos para as visualizações a desenvolver, que deverão permitir representar:

1. Número total de habitantes por zona geográfica, utilizando as NUTS2 e NUTS3.
2. Número de trabalhadores residentes por cada zona geográfica, utilizando as NUTS2 e NUTS3.
3. Deslocações das pessoas entre residência e trabalho.
4. Residência: a atual, há 1 ano e há 5 anos.

Para responder às necessidades do projeto foi criado o modelo lógico da base de dados que está representado na figura 3.2. O modelo é composto pelas seguintes tabelas: dados, nuts02, nuts03, alfabetismo, meio vida e sexo.

¹⁶ <https://www.postgresql.org/> (consultado a 26-09-2016)

¹⁷ <http://www.bulzip.com/products/a2p/info.php> (consultado a 26-09-2016)

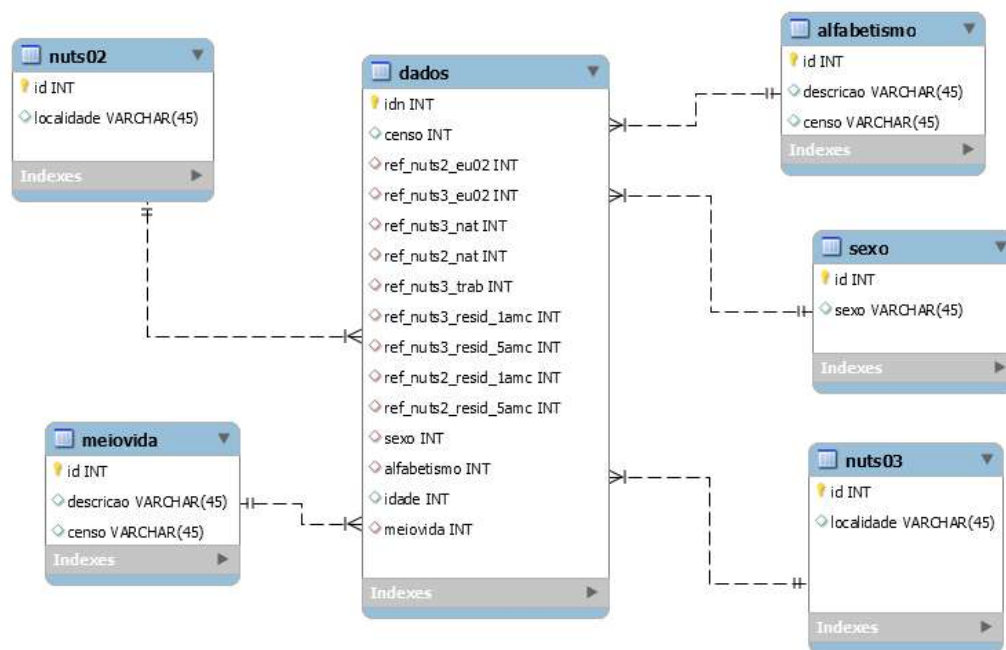


Figura 3.2 Modelo lógico da base de dados

Durante o desenvolvimento ficou definido que cada região tivesse uma cor associada. Para tal, foi adicionada uma nova coluna em cada tabela das NUTS. A inserção foi realizada de forma manual na base de dados. Para inserir esta informação foi necessário utilizar uma ferramenta de edição de imagem, que permitisse obter o código das cores de cada região. Foi criado um ficheiro em Excel para guardar as cores, que pode ser utilizado mais tarde.

4. Visualizações Desenvolvidas

Neste capítulo, após uma curta introdução serão apresentados e descritos os requisitos funcionais, as visualizações geradas, os controlos, as interações e os resultados finais.

4.1 Introdução

Com o modelo de dados obtido anteriormente é possível gerar as visualizações. Como exemplo, a figura 4.1, ilustra a representação do fluxo migratório entre duas regiões A e B.

Na região A um total de 3000 pessoas, 2500 pessoas mantiveram-se na mesma região de residência e de trabalho e 500 pessoas deslocaram-se para a região B. Já na região B 2000 pessoas mantiveram-se no mesmo local de residência e de trabalho.

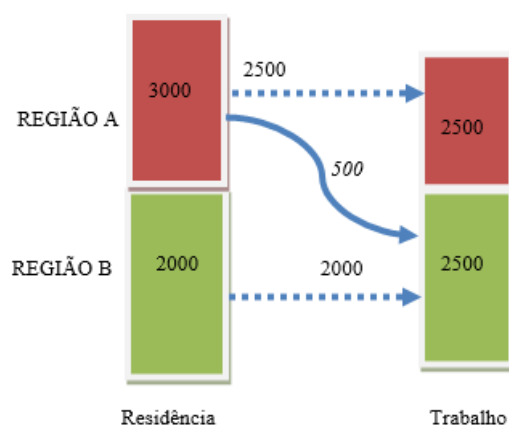


Figura 4.1 Exemplo de deslocação da residência para o trabalho

4.2 Objetivos e Requisitos Funcionais

As visualizações criadas neste trabalho têm como objetivo fundamental mostrar os fluxos migratórios em Portugal apresentando a informação de forma organizada.

Foram definidos alguns requisitos funcionais para este projeto. A aplicação deve assim permitir:

- A utilização de filtros para selecionar os dados a serem apresentados.
- A seleção do tipo de informação que se pretende visualizar; visualizar fluxo o local de residência para local de trabalho, bem como o local de residência atual, 1 ano antes e 5 anos antes.
- Representar os dados consoante o tipo de visualização selecionada seja o diagrama Cordal ou de Sankey.

4.3 Visualizações Geradas

Após implementação do modelo da base de dados, procedeu-se ao desenvolvimento das visualizações. As visualizações implementadas neste projeto são o Diagrama de Sankey e o Diagrama Cordal.

4.4.1 Diagrama de Sankey

A figura 4.2 apresenta uma visualização utilizando o diagrama de Sankey.

Para gerar o diagrama de Sankey recorreu-se a um *layout* já predefinido por Mike Bostock¹⁸. O diagrama de Sankey permite visualizar a magnitude dos fluxos das migrações. Na representação é possível visualizar o fluxo de saídas e de entradas de pessoas.

O utilizador, consegue visualizar fluxos migratórios no sentido de residência para o trabalho, a nível das NUTS3 e NUTS2 e pode ainda visualizar relações de residência que são:

- residência atual e há 1 ano antes
- residência atual e há 5 anos antes
- residência atual, há 5 anos antes e há 1 ano antes

¹⁸ <https://bost.ocks.org/mike/sankey/> (consultado a 04-10-2016)

Durante o desenvolvimento, decidiu-se também visualizar informação apenas de uma região seleccionada, clicando sobre uma determinada região obtém-se os destinos, como é possível verificar na figura 4.3.

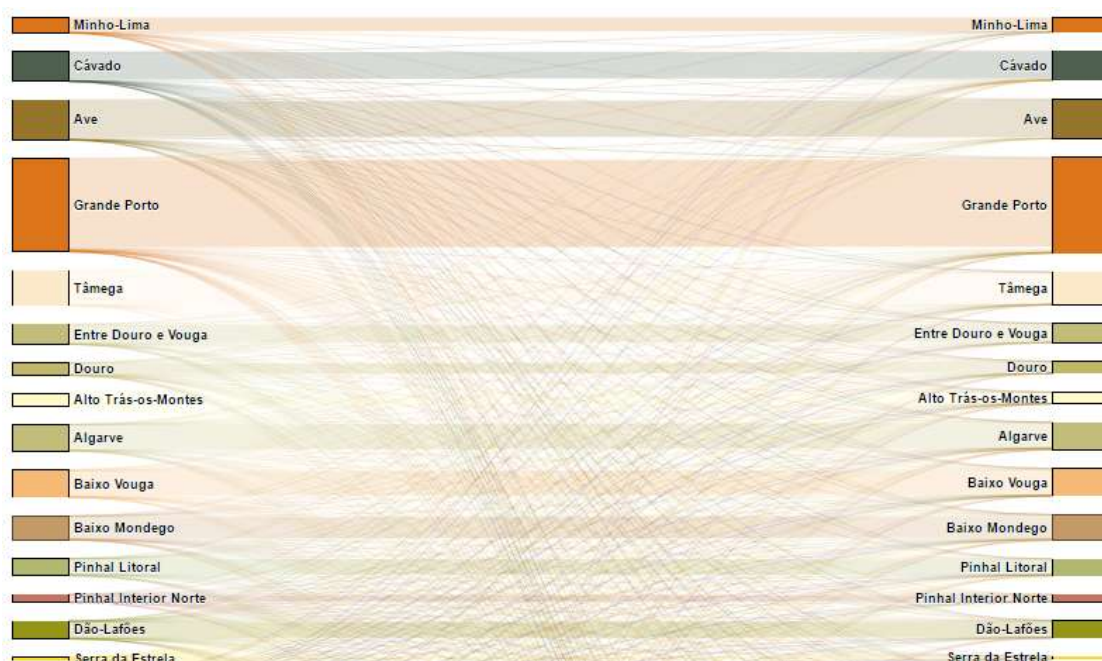


Figura 4.2 Diagrama de Sankey que representa as deslocações das pessoas da residência para o trabalho

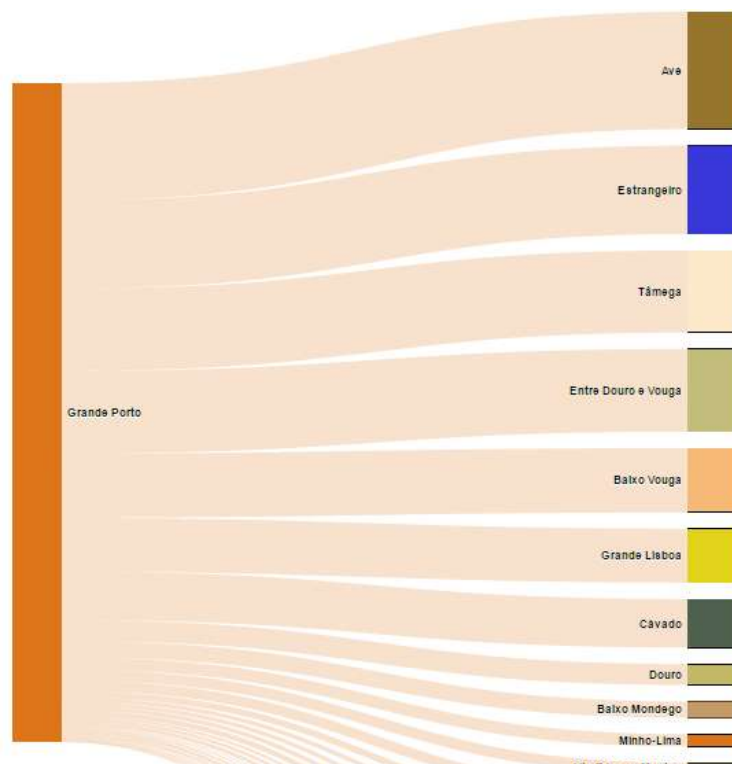


Figura 4.3 Diagrama de Sankey de uma zona residência para trabalho representando as deslocações para a região do Grande Porto

Na figura 4.4, é feita uma ilustração do fluxo migratório atendendo aos três instantes temporais de residência. A primeira coluna representa a região de residência atual, a segunda a residência de há um ano e a terceira a residência de há 5 anos.

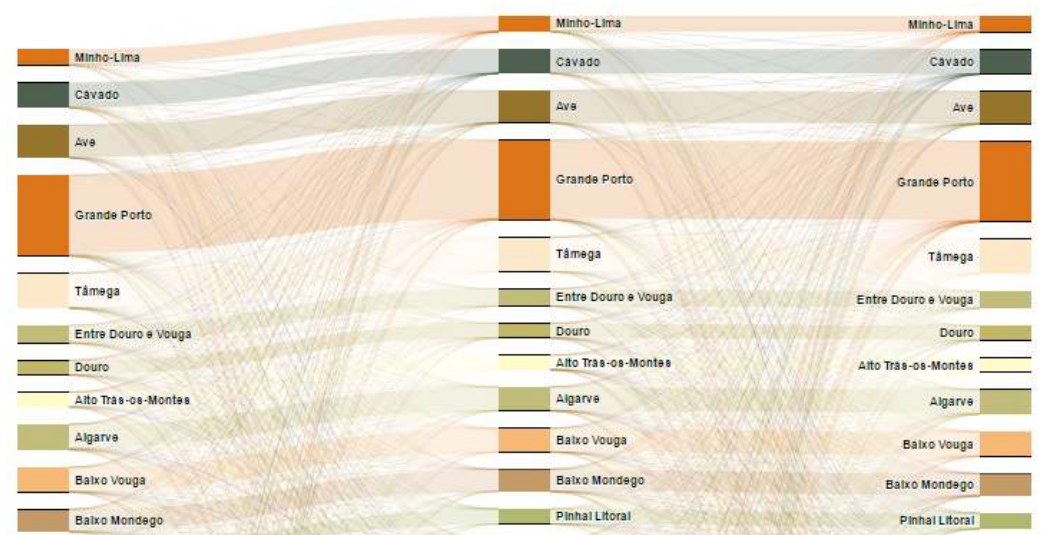


Figura 4.4 Diagrama de Sankey que mostra as alterações de residência relativamente ao ano de 2001(atual, há 1 ano e há 5 anos)

4.4.2 Diagrama Cordal

O diagrama cordal foi uma das representações desenvolvidas neste projeto, nesta visualização foi utilizado um *layout* D3.JS já predefinido (*Chord diagram*¹⁹). Esta visualização é uma das mais difíceis de implementar porque é necessário ter uma matriz com um formato predefinido e uma estrutura de dados complementar com nomes, cores e outros atributos que vão ser utilizados para apresentar a informação dos arcos.

O diagrama cordal desenvolvido foi inspirado no exemplo da secção 2.4. Neste tipo de representação os arcos representam o fluxo, isto é, a quantidade de indivíduos que se movimentam de uma região para outra. A figura 4.5 representa o fluxo migratório entre residência atual e há um ano.

Neste diagrama o utilizador consegue obter uma informação com mais detalhe e de forma mais rápida do que com o diagrama de Sankey. Apenas é necessário seleccionar uma região para visualizar a informação quanto às deslocações em ambos os sentidos. No diagrama de Sankey, o utilizador deverá seleccionar duas regiões (origem e destino) para obter informação detalhada.

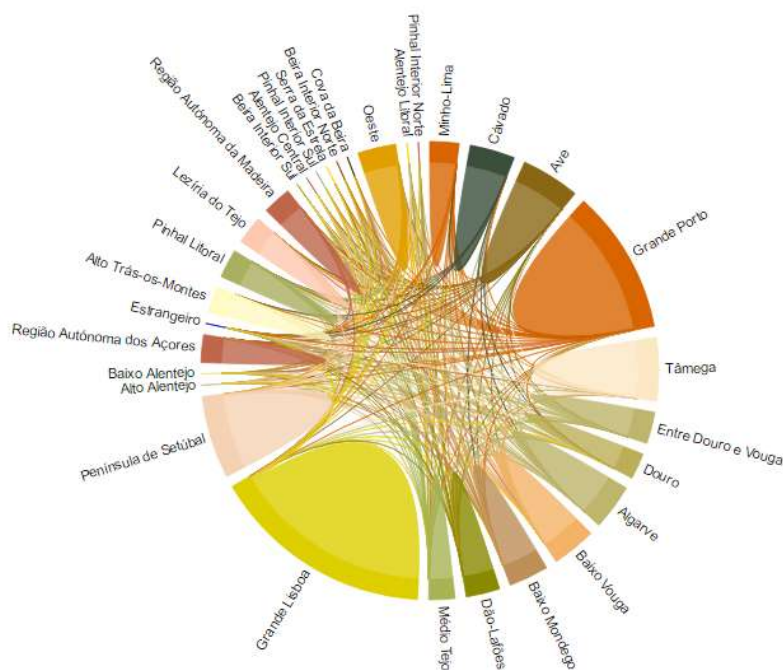


Figura 4.5 Diagrama Cordal representando a residência atual e há 1 antes utilizando as NUTS3

¹⁹ <https://bl.ocks.org/mbostock/4062006> (consultado a 07-10-2016)

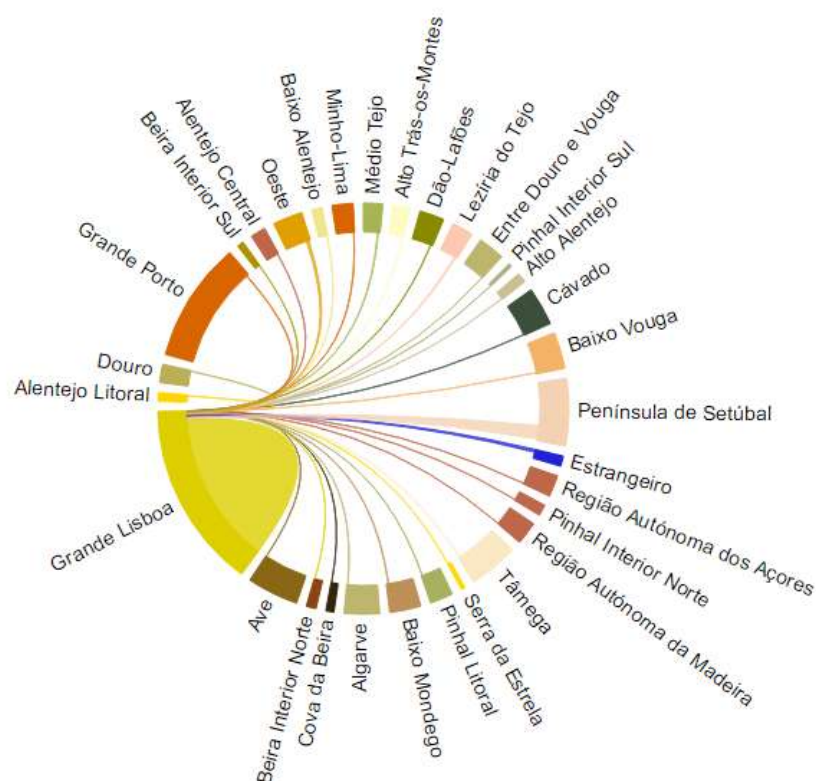


Figura 4.7 Espessura dos arcos associados à região da Grande Lisboa.

Visualização da residência atual para o trabalho

Além disto, através da espessura dos arcos é possível ver de onde é que entraram ou saíram mais pessoas, como já foi referido anteriormente; na figura 4.7, é possível verificar que na Península de Setúbal houve um maior fluxo migratório para Lisboa do que no sentido inverso.

4.4 Operações e Interação

As visualizações implementadas estão acompanhadas por um painel (figura 4.8) que permite fazer uma filtragem dos dados, para gerar uma nova visualização. Antes de gerar qualquer visualização, o utilizador deve também escolher o tipo de gráfico que deseja visualizar.



The image shows a web interface for a Sankey diagram. At the top, there is a breadcrumb trail: "Home / Sankey". Below this, there are several filter controls. On the left, there are six dropdown menus: "Ano:" (set to 2001), "Sexo:" (empty), "Meio Vida:" (empty), "Faixa Etária" (empty), and "Alfabeto:" (empty). On the right, under the heading "Tipo de Gráfico", there are three checkboxes: "Trabalho" (unchecked), "Residência" (checked), and "Nut 2" (unchecked). Below these is a dropdown menu set to "1 ano antes". At the bottom right of the filter section, there are two checkboxes: "Nut 2" (unchecked) and "Nut 3" (checked).

Figura 4.8 Filtragem de dados

Na figura 4.9 é representada a utilização da filtragem de dados. Neste exemplo é feita a seleção do ano 2001 e do sexo feminino. Por omissão o sistema utiliza o ano 2001. Para visualizar o ano de 2011 o utilizador tem de fazer um novo pedido.

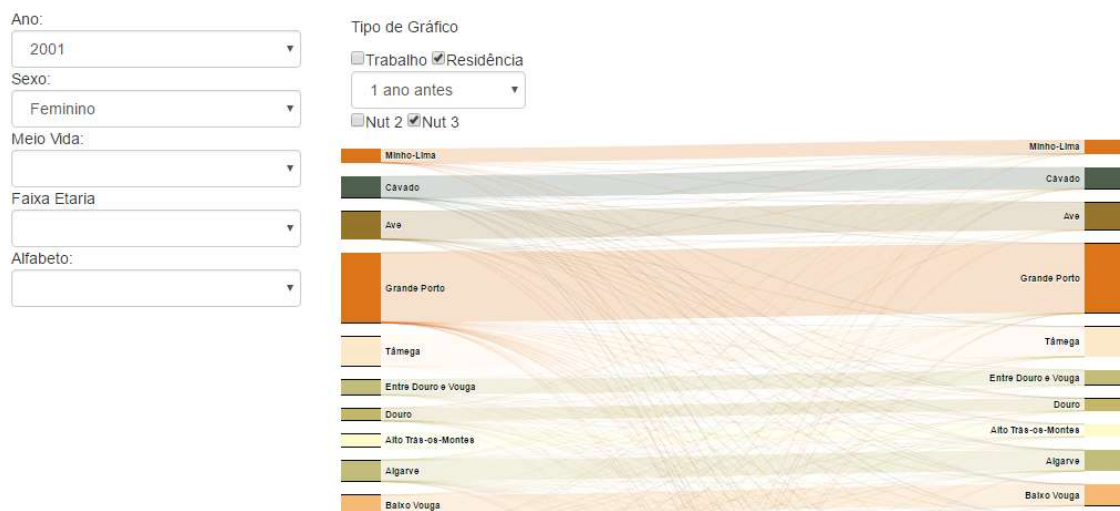


Figura 4.9 Digrama de Sankey com filtragem de dados

Uma das interações com a visualização (Figura 4.10) é a que permite obter informação sobre a deslocação das pessoas utilizando uma *tooltip*. Nesta imagem é possível ver informação adicional das deslocações das pessoas através da *tooltip*. Neste caso em particular não existe deslocação, porque a **origem** é o Grande Porto e o **destino** é o Grande Porto. É possível visualizar a percentagem de habitantes que não se deslocaram e o seu número.

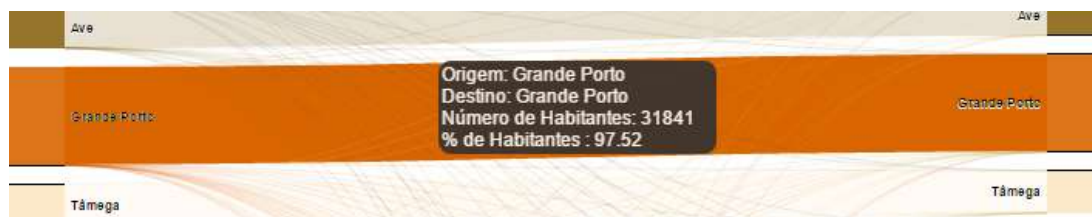


Figura 4.10 Tooltip de informação no diagrama de Sankey

A *tooltip* do diagrama cordal, na figura 4.11, permite obter um conjunto de informações mais completo e com mais clareza, isto é, de forma mais intuitiva para o utilizador, do que com o diagrama de Sankey.

Esta *tooltip* permite obter a percentagem daqueles que vão de uma região para outra e vice versa e permite ainda obter o *Net Flow*²¹, que é a diferença entre os que saem e entram.

²¹ https://en.wikipedia.org/wiki/Net_migration_rate (consultado a 07-10-2016)

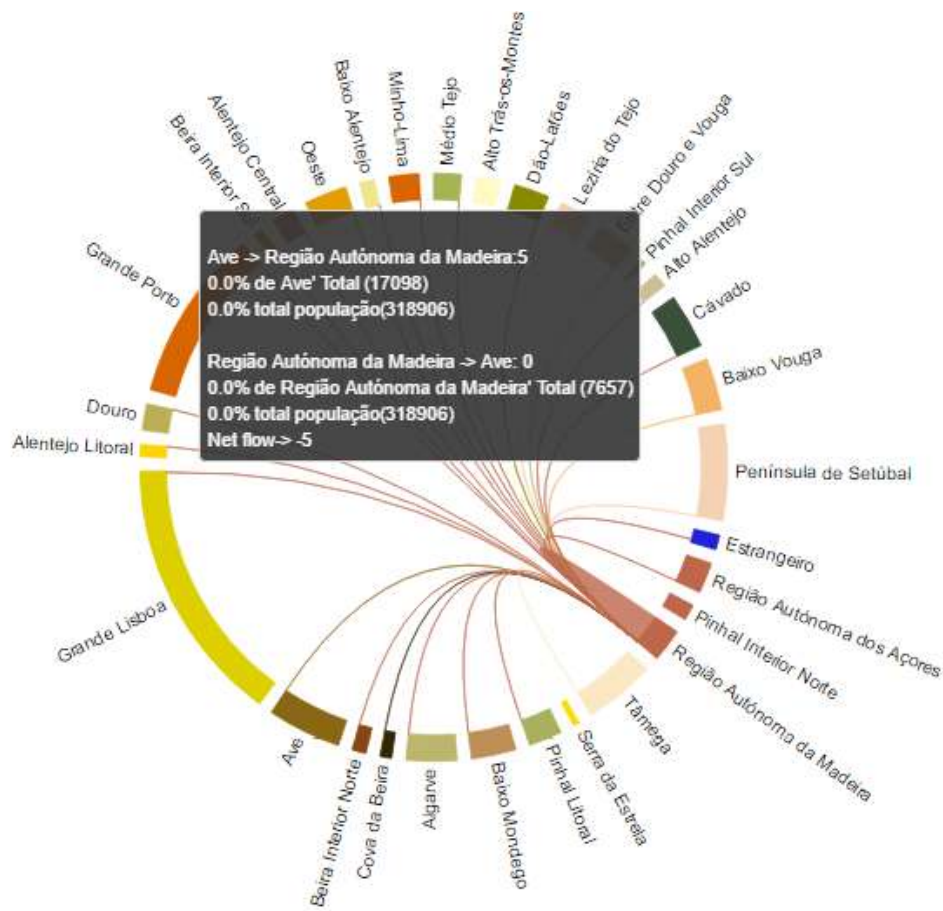


Figura 4.11 Tooltip do diagrama cordal

4.5 Resultados

Através das visualizações implementadas é possível obter um conjunto de informações sobre o fluxo migratório.

Podemos concluir que as zonas com maior fluxo migratório desta amostra são o Grande Porto e a Grande Lisboa.

São apresentados de seguida alguns resultados da amostra de dados para o ano de 2001 utilizada neste projeto.

As regiões que têm mais pessoas a sair do seu local de residência para o local de trabalho são: a Lezíria do Tejo com cerca de 16.2% e o Tâmega com 16.7%.

A Região com o menor número de pessoas a residirem na altura do recenseamento é o Alto de Trás-os-Montes.

Ao longo do período em análise, a Grande Lisboa manteve-se a região com mais população. A população residente na Grande Lisboa no ano de 2001 era de 97059 pessoas e em 2011 era de 97345 pessoas. Entre os anos 2001 e 2011 a população sofreu um aumento de 0.3% aplicando a fórmula da taxa de crescimento real²².

²² <http://pt.wikihow.com/Calcular-Taxa-de-Crescimento> (consultado a 07-10-2016)

5. Plataforma

As visualizações desenvolvidas foram integradas numa plataforma simples e intuitiva apresentada na figura 5.1.

Cada visualização tem um texto introdutório a explicar a maneira como é que o utilizador interage com essa visualização.

A plataforma contém uma pequena descrição sobre os dados utilizados neste projeto.

A plataforma é composta, assim, pelas páginas: página inicial, página com diagrama de Sankey e a página com o Diagrama Cordal. Na página inicial, o utilizador pode escolher qual a visualização que pretende ver.

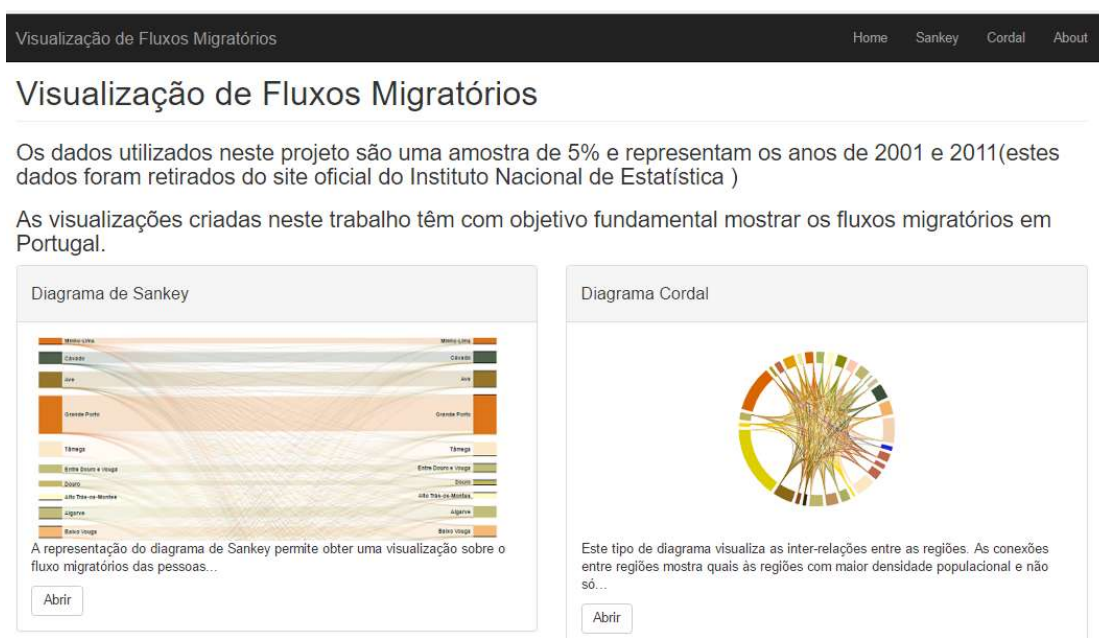


Figura 5.1 Plataforma de visualização de fluxos migratórios

5.1 Atores

A plataforma desenvolvida tem como público-alvo toda a comunidade que pretenda analisar fluxos migratórios em Portugal.

Foram identificados dois atores

1. Público em Geral - o utilizador que pretende visualizar os dados.
2. Sistema – que trata os dados e gera as visualizações seleccionadas.

5.2 Modelo de Casos de Uso

Com base na identificação dos utilizadores do sistema procedeu-se à criação do modelo de casos de uso conforme ilustrado na figura 5.2.

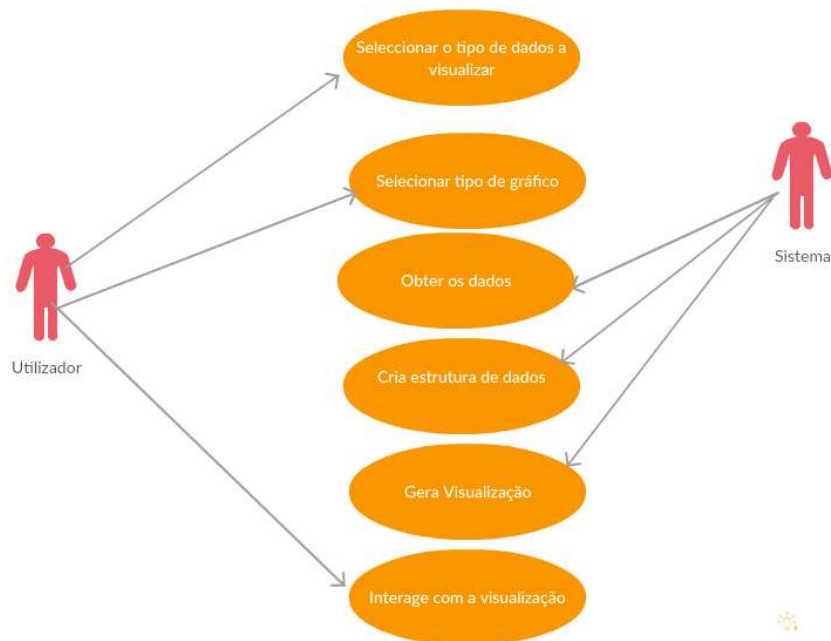


Figura 5.2 Modelo de casos de uso

Os casos de uso da plataforma representados figura 5.2 são os seguintes:

1. O utilizador selecciona os dados fornecidos que pretende visualizar, por omissão o ano seleccionado é 2001.
2. O utilizador escolhe qual o tipo de representação que pretende visualizar.
3. Tendo em conta as escolhas do utilizador o sistema obtém os dados.

5. O sistema depois de obter os dados representa-os num formato específico.
4. Com os dados estruturados é gerada visualização consoante a escolha do utilizador.
5. O utilizador pode interagir com a visualização e pode comparar dados entre as várias regiões.

5.3 Estrutura do Sistema

Tendo em consideração os casos de uso definidos anteriormente e a plataforma desenvolvida, foi desenhada a estrutura do sistema utilizando a estrutura MVC (*Model View Controller*) ²³ representada na figura 5.3. Esta estrutura é composta por 3 partes: controlador, modelo e vista.

O modelo é utilizado para interagir e armazenar os dados, como por exemplo, realizando consultas à base de dados e validação dos dados.

A vista consiste na representação dos dados.

O controlador faz a ligação entre o browser, vista e o modelo, recebe as informações do browser, envia-as para o modelo, recebe as informações do modelo e converte-as num formato para a vista.

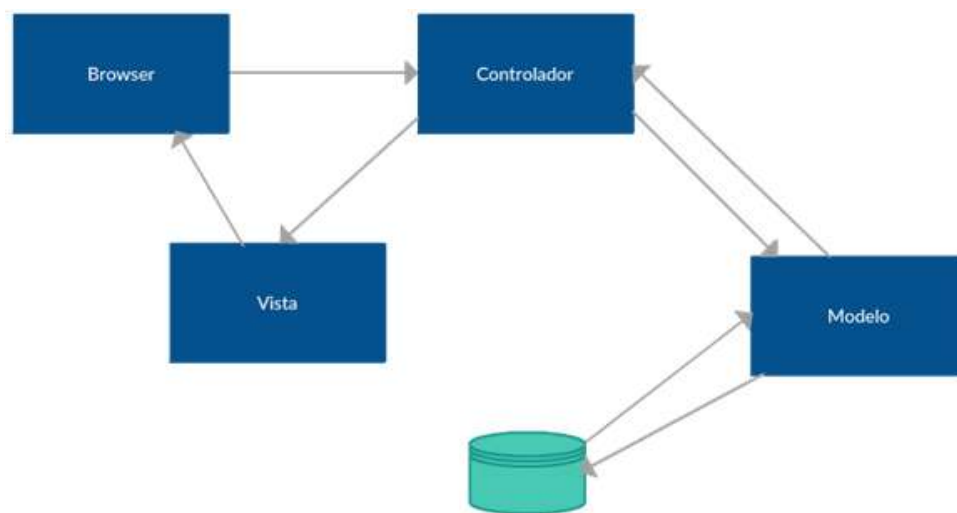


Figura 5.3 Modelo MVC

²³ <https://msdn.microsoft.com/en-us/library/ff649643.aspx> (consultado a 09-10-2016)

Com o modelo, descrito anteriormente, foi gerado um diagrama de componentes para representar a estrutura do sistema. O diagrama é composto por três componentes: aplicação do servidor, cliente e ainda pela base de dados. Na figura 5.4 está representado o diagrama de componentes.

A aplicação do cliente é o ambiente gráfico em que o utilizador pode interagir com a plataforma escolhendo que tipo de visualização pretende ver e que filtrações deseja aplicar.

O ambiente gráfico foi desenvolvido utilizando o HTML, *Bootstrap* para os estilos da plataforma e a utilização das bibliotecas *Javascript JQuery* e *D3.JS*, para tratamento de informação e criação das visualizações.

O servidor é composto pelo controlador e pelo modelo. Os pedidos realizados pelo cliente, podem ser feitos em *POST* ou *GET*, dependendo do tipo de visualização que o utilizador pretender. No diagrama cordal, os pedidos ao controlador são realizados em *GET* e no de Sankey em *POST*.

O controlador valida os parâmetros enviados pela *interface* gráfica e chama a função pretendida do modelo, esta executa uma *query* à base de dados, usando a biblioteca PDO (PHP Data Objects), e, de seguida, devolve os resultados ao controlador. Este processa a informação que será enviada ao cliente, em formato JSON ou CSV. A vista processa os dados recebidos e gera a visualização correspondente.

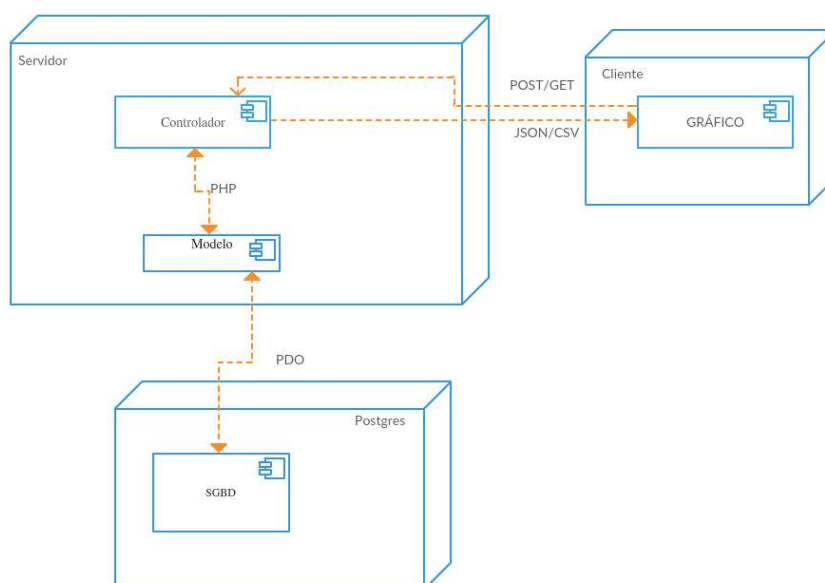


Figura 5.4 Diagrama de componentes

5.4 Estrutura de Ficheiros

São apresentados os formatos dos dados que servem de base às visualizações. A estrutura de dados utilizada no diagrama de Sankey está representada na Figura 5.5. Esta estrutura de dados está no formato JSON e será utilizada pelo D3 para a criação da visualização, em que os *links* representam as ligações das origens e dos destinos e ainda contém a seguinte informação: número total de pessoas que se deslocaram, o número total de cada origem e percentagem daqueles que se deslocaram. Os *nodes* servem para identificar as ligações entre as origens (*source*) e os destinos (*target*), definir a sua posição através do “xpos” em que o valor zero corresponde à primeira coluna e o um à segunda coluna do diagrama de Sankey.

```
"links": [
  {
    "id": "111",
    "color": "#d96500\r\n\r",
    "source": "Minho-Lima ",
    "target": "Minho-Lima",
    "id2": "111",
    "value": "6029",
    "total": "6982",
    "percentagem": "86.35"
  },
  {
    "id": "111",
    "color": "#d96500\r\n\r",
    "source": "Minho-Lima ",
    "target": "Cávado",
    "id2": "112",
    "value": "252",
    "total": "6982",
    "percentagem": "3.61"
  },
  {
    "id": "111",
    "color": "#d96500\r\n\r",
    "source": "Minho-Lima ",
    "target": "Ave",
    "id2": "113",
    "value": "56",
    "total": "6982",
    "percentagem": "0.80"
  },
  {
    "id": "111",
    "color": "#d96500\r\n\r",
    "source": "Minho-Lima ",
    "target": "Grande Porto",
    "id2": "114",
    "value": "276",
    "total": "6982",
    "percentagem": "3.95"
  }
],
"nodes": [
  {
    "name": "Minho-Lima ",
    "color": "#d96500\r\n\r",
    "xpos": "0"
  },
  {
    "name": "Minho-Lima",
    "color": "#d96500\r\n\r",
    "xpos": "1"
  },
  {
    "name": "Cávado ",
    "color": "#3b4f3b\r\n\r",
    "xpos": "0"
  },
  {
    "name": "Cávado",
    "color": "#3b4f3b\r\n\r",
    "xpos": "1"
  },
  {
    "name": "Ave ",
    "color": "#8a6715\r\n\r",
    "xpos": "0"
  },
  {
    "name": "Ave",
    "color": "#8a6715\r\n\r",
    "xpos": "1"
  }
],
```

Figura 5.5 Links e nodes do diagrama de Sankey

Na figura 5.6 e 5.7 são apresentadas as estruturas de dados que são utilizadas no D3 para a criação do diagrama cordal.

Na figura 5.6 está representada a matriz no formato JSON, em que os valores das células indicam a número total de pessoas que se deslocaram de um lado para o outro. Cada linha corresponde a uma origem e a coluna um destino.

```
[ [5396,1,3,143,1,9,43,5,19,15,78,3,14,2,1,0,155,52,20,6,10,3,326,0,6,24,15,15,24,5,0],
  [6,5122,5,4,2,0,3,42,10,10,100,5,1,6,30,0,6,19,12,2,12,23,70,0,94,147,2,4,3,0,1],
  [14,7,7191,8,11,0,7,11,124,11,177,6,23,10,10,31,18,139,10,16,32,4,146,5,41,88,5,8,18,2,3],
  [152,2,7,5991,1,0,12,2,5,73,74,5,3,6,1,0,32,17,19,6,7,1,559,4,4,9,8,31,114,4,0],
  [2,9,16,2,7799,0,4,16,248,2,60,1,2,1,21,0,5,47,6,8,3,12,22,1,18,534,4,3,3,4,2],
  [29,0,1,9,0,901,3,2,2,3,19,1,12,0,1,0,1,7,0,5,1,1,35,1,0,0,45,1,4,2,0],
  [28,0,1,14,0,1,2940,0,3,22,18,0,0,0,1,2,4,6,16,5,0,0,126,5,2,2,22,92,3,11,0],
  [5,29,7,4,3,0,2,11187,28,3,134,9,0,6,27,0,11,24,10,9,12,434,83,0,30,411,5,2,4,5,220],
  [15,11,56,12,328,3,6,15,10701,8,98,3,8,6,7,0,20,304,13,15,22,6,98,0,18,295,11,6,5,5,9],
  [26,4,2,48,2,1,17,5,10,16432,244,14,1,10,4,0,10,32,74,3,15,4,5109,52,11,31,12,68,37,35,2],
  [55,46,77,69,60,2,18,65,106,239,310,31,18,66,53,12,90,59,335,14,22,63,1071,28,45,202,15,46,1],
  [3,1,3,6,0,0,1,5,9,11,31,6838,0,15,1,0,1,16,8,1,2,1,121,0,3,21,1,1,6,3,1],
  [28,0,32,5,2,6,4,4,14,5,40,0,2985,0,0,18,49,345,10,15,17,0,85,0,1,13,13,0,14,3,0],
  [3,3,2,4,0,0,2,10,6,9,120,7,0,7291,3,0,4,28,6,2,2,0,83,2,4,48,4,3,2,6,3],
  [8,53,20,8,65,0,7,72,38,6,185,16,2,9,13793,1,19,30,25,4,25,347,109,3,132,1561,5,7,6,10,12],
  [2,1,60,2,1,0,1,1,9,3,27,0,22,0,0,1087,4,17,2,8,34,1,20,0,1,9,1,0,0,1,0],
  [176,1,7,41,1,0,3,3,20,18,48,2,16,2,5,1,6825,189,14,6,5,1,139,1,5,21,8,7,119,5,1],
  [37,3,72,20,7,7,4,13,331,15,74,16,109,6,7,5,169,9086,25,11,13,4,197,3,9,78,14,14,20,3,0],
  [5,3,0,6,2,0,7,3,5,33,80,4,6,2,6,1,11,16,11202,0,2,4,207,9,4,8,1,23,6,50,7],
  [5,2,11,3,0,1,6,1,3,5,25,1,7,3,0,3,5,23,10,2325,44,0,54,1,2,10,71,4,5,1,0],
  [4,4,18,4,6,0,4,3,15,3,36,5,0,1,3,21,4,49,6,59,2871,1,51,3,21,20,17,3,5,0,1],
  [4,32,3,4,7,1,5,423,19,2,139,4,0,5,293,1,9,26,18,7,4,15250,51,0,30,695,5,6,3,6,46],
  [74,18,29,292,7,11,38,19,38,945,612,32,15,44,21,6,47,81,198,13,60,15,58761,37,25,98,15,80,42],
  [2,1,1,1,0,1,6,1,1,71,36,1,0,1,0,0,5,6,58,1,0,0,94,2395,1,4,0,30,3,60,0],
  [9,118,91,9,2,1,5,10,29,4,163,3,5,2,57,0,5,26,16,11,37,5,89,1,5120,218,7,10,6,1,4],
  [19,49,29,9,370,6,6,217,288,16,396,24,5,16,368,1,19,73,46,18,12,526,241,4,94,37137,10,12,14],
  [9,1,0,6,1,9,12,2,6,9,15,0,5,3,2,1,3,20,10,31,8,0,53,0,0,2,1935,4,1,1,0],
  [13,1,2,16,0,0,66,0,4,67,17,1,0,1,0,0,1,7,24,6,1,0,129,17,0,6,3,4588,8,64,0],
  [25,3,8,122,0,0,8,3,11,48,75,2,6,1,1,1,203,39,23,2,8,0,902,9,4,16,2,19,8302,9,0],
  [3,0,0,0,1,0,8,0,2,23,44,1,0,0,1,0,4,4,78,1,0,1,101,26,2,2,0,53,3,3108,0],
  [6,16,2,2,8,0,3,252,19,6,107,4,0,3,6,0,7,18,15,6,3,56,96,1,25,276,2,11,0,3,6029] ]
```

Figura 5.6 Exemplo de uma matriz do diagrama cordal

Na figura 5.7 estão apresentados os atributos no formato CSV que vão ser utilizados para identificar as ligações entre os arcos e a cor da região no diagrama Cordal.

```
name,color
"Alentejo Central",#bf664b
"Grande Porto",#d96588
"Dão-Lafões",#8a8a88
"Região Autónoma da Madeira",#bf664b
"Pinhal Interior Norte",#bf664b
"Grande Lisboa",#decf88
"Médio Tejo",#a7b555
"Baixo Vouga",#f5b367
"Baixo Alentejo",#f8e68d
Oeste,#e39f88
"Península de Setúbal",#f5d2b3
Mínho-Lima,#d96588
Douro,#bda5f5
"Cova da Beira",#332687
Estrangeiro,#2222d6
"Pinhal Litoral",#a9b85f
"Baixo Mondego",#bf8f58
"Região Autónoma dos Açores",#bf664b
Algarve,#bdb66c
"Beira Interior Sul",#ab9488
```

Figura 5.7 Excerto dos atributos da matriz do diagrama cordal

6. Avaliação

A utilização de métodos de avaliação das visualizações criadas é cada vez mais importante na área de visualização de informação [19].

Neste capítulo, pretende-se descrever as avaliações heurísticas realizadas à plataforma e apresentar os resultados obtidos.

6.1 Avaliação Heurística

A avaliação heurística é um método simples para analisar a usabilidade de uma interface, possibilitando, assim, um *feedback* rápido sobre os problemas de usabilidade da plataforma [19].

A avaliação heurística deve ser realizada individualmente por cada avaliador. Este procedimento é importante para garantir avaliações independentes e únicas. As respostas devem ser feitas num único documento e individualmente.

Para a realização de testes devem ser utilizados 3-5 avaliadores para encontrar problemas de usabilidade [20].

No caso deste projeto, a realização desta avaliação é para detetar anomalias para serem corrigidas no futuro.

Na avaliação da plataforma foram consideradas as heurísticas de Jakob Nielsen e de Zuk et al., que são descritas a seguir.

Heurísticas de Jakob Nielsen

As heurísticas de Nielsen estão agrupadas em 10 princípios, que são utilizados para avaliar a usabilidade da aplicação *web* [21] e têm como objetivo prevenir erros que podem prejudicar a utilização por parte de um utilizador.

Lista de Heurísticas:

1. Visibilidade do estado sistema

Fornecer ao utilizador *feedback* sobre o estado do sistema.

Exemplo: se uma página demorar muito tempo a carregar, deve exibir uma barra de progresso para que os utilizadores saibam que têm de esperar.

2. Não utilizar termos complicados

O sistema deve falar a linguagem dos utilizadores, com palavras, frases e conceitos familiares. As informações devem de ser organizadas de forma natural e lógica.

Exemplo: ao criar um site para crianças, devem-se utilizar termos com os quais estão familiarizados.

3. O utilizador tem controlo e liberdade

Permitir que o utilizador possa desfazer ou refazer uma determinada ação no sistema e voltar ao ponto anterior, caso se tenha enganado ou se estiver perdido.

4. Consistência

Nunca identificar a mesma operação com ícones ou palavras diferentes. A utilização de uma linguagem visual consistente facilita a compreensão por parte do utilizador.

5. Prevenção de erros

Melhor do que uma boa mensagem de erro é um design cuidadoso que possa evitar esses erros. Ao remover dados do sistema deve ser sempre requerida a confirmação do utilizador antes de apagar os dados.

6. Reconhecer em vez de lembrar

O sistema deve mostrar os elementos de diálogo e permitir que o utilizador faça as suas escolhas, sem a necessidade de lembrar um comando específico.

7. Desenho e *design* simples

Apresentar exatamente a informação que o utilizador precisa, no momento, nem mais nem menos.

8. Flexibilidade e Eficiência

Permitir o uso de atalhos para utilizadores experientes executarem as operações mais rapidamente.

9. Ajudar os utilizadores a reconhecer, diagnosticar e reparar erros

As mensagens de erro do sistema devem ser apresentadas de forma simples, indicar com precisão o problema e sugerir uma solução.

10. Ajuda e documentação

O ideal é que o sistema possa ser usado sem documentação, mas pode ser necessário oferecer ajuda, por isso, deve disponibilizar documentação.

Heurísticas de Zuk et al.

As heurísticas de Zuk et al. são utilizadas para avaliação da usabilidade de visualizações. Estas heurísticas têm como objetivo de prevenir erros [22].

Lista das heurísticas:

1. Confirmar que a variável visual tem comprimento suficiente.
2. A cor não deve induzir uma ordenação.
3. A perceção da cor varia com o tamanho do item colorido.
4. O contraste local afeta a perceção da cor e do cinzento.
5. Considerar as pessoas daltónicas.
6. Benefícios pré-atentivos aumentam com o campo de visão.
7. A avaliação quantitativa requer uma variação de tamanho ou posição
8. Ajustar os dados à dimensionalidade dos gráficos.
9. Colocar a maior quantidade de dados no menor espaço possível.
10. Disponibilizar vários níveis de detalhe.
11. Integrar texto onde for relevante.

6.2 Resultados

Atendendo às heurísticas anteriormente descritas, foi feita uma avaliação à plataforma, realizada por dois alunos de MSI, ambos do sexo masculino, com bastante conhecimento em informática. Para realizar esta avaliação cada elemento testou aplicação e foram registados alguns problemas encontrados. Com a realização da avaliação, pretende-se encontrar erros de usabilidade da plataforma e foram classificados na seguinte escala:

- 1-Não tem importância: não é necessário corrigir
- 2- Cosmético: não precisa de ser corrigido a menos que se disponha de tempo extra
- 3-Problema Menor: baixa prioridade à resolução deste problema
- 4-Problema Grave: deve ser dada alta prioridade à resolução deste problema

Alguns dos problemas encontrados:

- Pode existir um erro de ligação à base de dados e não é dado qualquer feedback ao utilizador.
 - Heurística Nielsen → Visibilidade do Estado do Sistema
 - Nível → 4
- Nomes que não estão corretamente escritos, como por exemplo: na filtragem de dados deve-se escrever alfabetização e não alfabeto.
 - Heurística Nielsen → Não utilizar termos complicados
 - Nível → 2
- No diagrama de Sankey, podem existir regiões sem ligação.
 - Heurística Nielsen → Desenho e design simples
 - Nível → 3
- Não é dada nenhuma mensagem ao utilizador se não existirem dados para gerar a visualização
 - Heurística Nielsen → Visibilidade do Estado do Sistema
 - Nível → 3
- O utilizador não sabe o que deve seleccionar em primeiro lugar para gerar a visualização
 - Heurística Nielsen → Desenho e design simples
 - Nível → 3

- A plataforma não suporta utilizadores daltónicos.
 - Heurística Zuk → Considerar as pessoas daltónicas
 - Nível → 4

Conclui-se que a aplicação cumpre as seguintes heurísticas:

- O sistema informa o utilizador de que está a processar informação através de um ícone de espera.
 - Heurística → Visibilidade do Estado do Sistema
- O utilizador é livre de escolher que visualização quer ver.
 - Heurística → O utilizador tem controlo e liberdade
- Aplicação tem uma interface simples.
 - Heurística → Desenho e design simples
- O utilizador pode voltar para o menu inicial.
 - Heurística → Utilizador tem controlo e liberdade
- Cada visualização tem um texto introdutório a explicar o seu funcionamento.
 - Heurística → Ajuda e Documentação
- O utilizador pode fazer filtragem de dados.
 - Heurística → O utilizador tem controlo e liberdade
- Aplicação permite obter informações sobre os fluxos migratórios através de *tooltips*.
 - Heurística → Flexibilidade e Eficiência
- No diagrama de Sankey o utilizador pode seleccionar uma região e ver apenas as ligações desta.
 - Heurística → Flexibilidade e Eficiência
- O utilizador pode voltar à visualização anterior, clicando sobre a região
 - Heurística → Flexibilidade e Eficiência
- É possível analisar as diferenças das quantidades de pessoas que entram e saem no digrama cordal e no de Sankey.
 - Heurística → A avaliação quantitativa requer uma variação de tamanho ou posição

- É apresentada informação contextual ao utilizador quando passa o cursor por cima de cada ligação.
 - Heurística → Integrar texto onde for relevante

Após a avaliação de usabilidade ficou resolvido o problema de termos que não estão corretamente escritos, foi prontamente resolvido (trocado na filtragem "alfabeto" por "alfabetização").

7. Conclusão e Trabalho Futuro

Neste projeto foi desenvolvida uma plataforma que tem o objetivo de representar os dados dos fluxos migratórios dos censos de 2001 e 2011. As visualizações criadas permitem representar: fluxo migratório da residência para o trabalho, residência atual para residência há 1 ano antes e 5 anos antes.

Os dados fornecidos pelo INE não possuíam informação suficiente sobre os municípios, pois era interessante visualizar as quantidades de pessoas por município e ver o crescimento entre 2001 e 2011.

Para a realização deste projeto, foi fundamental alguma leitura sobre visualização de informação e sobre os censos. Também foi necessário realizar uma análise profunda aos dados facultados pelos censos assim como ler documentação sobre as visualizações a serem desenvolvidas.

O diagrama de Sankey representa o número de pessoas que foram de uma região para outra.

No digrama cordal os arcos representam o número das pessoas que saem e entram de uma região para outra e vice-versa, permite obter ainda o Net Flow.

Numa próxima iteração deste projeto pode-se melhorar a interatividade das visualizações como por exemplo: a criação de uma tabela complementar para representar as quantidades de destino e no sentido inverso de uma determinada região. Desta forma, o utilizador poderá ver melhor os dados existentes bem como identificar as relações entre as regiões.

Durante o desenvolvimento foram encontrados alguns problemas que podem ser corrigidos numa próxima fase do projeto, como por exemplo; falta de uma segunda

legenda nos gráficos a indicar o que se está a visualizar, isto é que fluxo está a ver. E também devem ser corrigidos os erros encontrados na avaliação heurística.

8. Referências

- [1] M. Bostock, V. Ogievetsky e J. Heer, “A tour through the Visualization zoo,” *Commun. ACM*, vol. 53, pp. 59--67, 2010.
- [2] L. Shixia, C. Weiwei, W. Yingcai e L. Mengchen, “A survey on information visualization: recent advances and challenges,” *Vis. Comput.*, pp. 1373-1393, dec, 2014.
- [3] C. Chen, “Information visualization,” *Wiley Interdisciplinary Reviews: Computational Statistics*, pp. 387-403, 2010.
- [4] A. Kerren, J. T. Stasko, J.-D. Fekete e C. North, *Information Visualization*, Berlin, Heidelberg: Springer-Verlag, 2008, pp. 8-13.
- [5] W. Colin, *Information Visualization*, 2 ed., Morgan Kaufmann, 2004.
- [6] R. Mazza, *Introduction to Information Visualization*, 1 ed., Switzerland: Springer-Verlag London, 2009, pp. 20-26.
- [7] P. Riehmann, M. Hanfler e B. Froehlich, “Interactive Sankey diagrams,” *IEEE Symposium on Information Visualization*, pp. 233-240, 2005.
- [8] “Infovis-wiki,” [Online]. Available: http://www.infovis-wiki.net/index.php?title=Visualization_Pipeline&oldid=14045. [Acedido em 21 09 2016].
- [9] B. Prantik, P. Abhisek e B. Paritosh, “Application of (0, 1)-Matrix in Determination of Graphic Realization of Non-Increasing Positive Integer Sequence,” *In Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015) (ICARCSET '15)*, pp. 6:1-6:5, 2015.
- [10] N. Sander, G. J. Abel, R. Bauer e J. Schmidt, “Visualising Migration Flow Data,” Viena-Austria, 2014.
- [11] T. Munzner, *Visualization Analysis and Design*, 1º ed., CRC Press, 2014.

- [12] J. Heer e B. Shneiderman, “Interactive dynamics for visual analysis.,” vol. 54, pp. 45-54, 2012.
- [13] Google, “Google Charts,” [Online]. Available: <https://developers.google.com/chart/interactive/docs/gallery/geochart>. [Acedido em 23 09 2016].
- [14] “Charts.js,” [Online]. Available: <http://www.chartjs.org/>. [Acedido em 23 09 2016].
- [15] “Flotcharts,” [Online]. Available: <http://www.flotcharts.org/>. [Acedido em 23 09 2016].
- [16] “Dc.js,” [Online]. Available: <https://dc-js.github.io/dc.js/>. [Acedido em 23 09 2016].
- [17] M. Bostock, V. Ogievetsky e J. Heer, “D3 Data-Driven Documents,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, pp. 2301-2309, December 2011.
- [18] INE, “Estatísticas Demográficas 2011,” Instituto Nacional de Estatística, IP, Lisboa, 2013.
- [19] F. Camilla e J. Jimmy, “An Heuristic Set for Evaluation in Information Visualization,” *Proceedings of the International Conference on Advanced Visual Interfaces*, vol. 10, pp. 199-206, 2010.
- [20] J. Mankoff, A. K. Dey, G. Hsieh, J. Kientz, S. Lederer e M. Ames, “Heuristic Evaluation of Ambient Display,” *Proceedings of the SIGCHI Conference on Human Factors in Computing System*, pp. 169-176, 2003.
- [21] J. Nielsen, “10 Usability Heuristics for User Interface Design,” [Online]. Available: <https://www.nngroup.com/articles/ten-usability-heuristics/>. [Acedido em 12 10 2016].
- [22] T. Zuk, L. Schlesier, P. Neumann, M. S. Hancock e S. Carpendale, “Heuristics for Information Visualization Evaluation,” *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, pp. 1-6, 2006.
- [23] “Epoch,” [Online]. Available: <https://epochjs.github.io/>. [Acedido em 23 09 2016].
- [24] J. M. C. Bastien e D. L. Scapin, “Preliminary Findings on the Effectiveness of Ergonomic Criteria for the Evaluation of Human-computer Interfaces,” *INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems*, pp. 187-188, 1993.
- [25] R. Mazza, “Introduction to Information,” University of Lugano, 2004.

Anexo A Organização dos ficheiros

Os ficheiros disponibilizados neste projeto estão organizados na seguinte forma:

- Uma diretoria com o nome “**dados**”, que contém o backup da base de dados utilizada no projeto com o nome de “**censos.backup**”.
- Uma diretoria com o nome de “**projeto**” que contém os ficheiros do projeto desenvolvido.

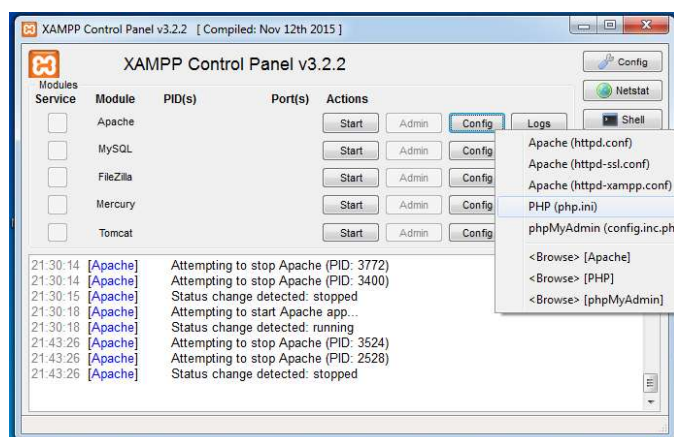
A pasta **projeto** contém as seguintes subdiretorias:

- **bd** com o ficheiro de ligação à base de dados com o nome `connect.php`.
- **d3** com as bibliotecas utilizadas neste projeto, `d3.js`, `jquery.js`, `mapper.js`, `sankey.js`, `underscore.js`, `queue.js`.
- **js** com ficheiros do bootstrap e jquery.
- **fonts** com tipo de letra.
- **css** com os css utilizados.
- **img** com as imagens utilizadas na plataforma.
- **processar** com os ficheiros `circular.js` e `sankey.js` estes dois ficheiros são utilizados para fazer a filtragem de dados e criar as visualizações.
- **querys** que contém os ficheiros `querys`, `Sankey`, `chord` que são utilizados para fazer os pedidos à base de dados e gerar a informação num formato requerido para a gerar as visualizações.
- **frontend** com as páginas `index.html`, `chord.html`, `sankey.html`.

Anexo B Manual de Instalação

De seguida são mostrados os passos necessários para instalação do projeto.

1. Instalar o PostgreSQL 9.4.10 ou superior, com a palavra-passe admin e a porta 5432 ou outra. Caso seja efetuada alguma alteração diferente da anterior deve ir ao ficheiro connect.php e alterar os mesmos.
2. Seguidamente é necessário criar uma base de dados com o nome “censos” no PostgreSQL e fazer o restauro do ficheiro censos.backup.
3. De seguida é necessário instalar um servidor. Neste caso, o servidor utilizado é o XAMPP. É necessário adicionar a linha `extension=php_pdo_pgsql.dll` no ficheiro `php.ini`. Por norma a extensão é incluída na instalação do XAMPP automaticamente, pode-se verificar se a extensão existe ou não em `C:\xampp\php\ext`.



4. Colocar a pasta do projeto em `C:\xampp\htdocs` e executar o Apache no XAMPP
5. Ir ao navegador e colocar o endereço `http://localhost:80/projeto/frontend/` para aceder à plataforma Web e escolher o tipo de visualização que quer ver.

